2013

PROBABILISTIC MODELING OF RAINFALL INDUCED LANDSLIDE HAZARD ASSESSMENT IN SAN JUAN LA LAGUNA, SOLOLÁ, GUATEMALA

Patrice F. Cobin
Michigan Technological University

PROBABILISTIC MODELING OF RAINFALL INDUCED LANDSLIDE HAZARD ASSESSMENT IN SAN JUAN LA LAGUNA, SOLOLÁ, GUATEMALA

By

Patrice F. Cobin

A REPORT

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

In Geology

MICHIGAN TECHNOLOGICAL UNIVERSITY

2013

This report has been approved in partial fulfillment of the requirements for the Degree of MASTER OF SCIENCE in Geology.

Department of Geological/Mining Engineering and Sciences

Report Co-Advisor: *Thomas Oommen*

Report Co-Advisor: *John S. Gierke*

Committee Member: *Michael Falkowski*

Department Chair: *John S. Gierke*

Table of Contents

Acknowledgements

A lot of uncertainty surrounds service as a Peace Corps volunteer. Only more uncertainty exists as a Peace Corps Masters International (PCMI) student wondering what your project will be. This project would not exist without the tremendous work the Geológos del Mundo (GM) undertook from 2010 to 2012, nor without Laura Nunez of GM who gladly shared with me all that they developed. A large portion of this research was possible through sharing of that nature and open source software. I will gladly share this study and data so the safety of those living around Lake Atitlán can be better served.

While the information shared with me played a huge part in allowing this project to exist, it would not have developed to the stage presented here without the guidance from John S. Gierke and Thomas Oommen. John, thank you for the time you took to visit me in my site, guiding me very early in my landslide interest. Those visits grounded me to my work, reminding me of my larger goal as a PCMI volunteer. Thomas, without your guidance during my time as a volunteer and during my time back at Michigan Tech the structure of this work would be both different and lacking. I would also like to acknowledge the generous support of the NSF and PIRE grant 0530109 that in part made this program and masters research financially feasible. I would also like to thank my third committee member Michael Falkowski for your time.

My Mom and Dad were the first champions of my wanderlust. I would also like to thank the Wills, Pagnanis, Cambriellos and Cobins for always encouraging me in my pursuits even when it took me away from them for years on end.

My time in Guatemala would not have been as enjoyable or successful without the many people I came to know in my site, other Peace Corps Volunteers (too many to name) and Peace Corps Staff. I would like to thank the team at Ati't Ala', specifically my counterparts who guided and supported: Raquel Ujpan, Maria José Mansilla, Brenda Noriega, Arturo Ujpan and Laura Wheelock. I would also like to thank the families of San Juan who made me part of their family for no other reason than the love they had to share, you are why Peace Corps volunteers continue to exist.

I would also like to thank all the support my friends here at Michigan Tech offered freely. Many questions would have gone unanswered without the help of Rudiger Escobar, Jay Welik and Stephanie Tubman. This time here finishing my masters would have been much poorer without the love and support of Evan Lucas.

Thank you to Kelly McLean and Amie Ledgerwood for all of your logistical help over the years.

And finally, Chispa, the friend I never expected to have; who has been unendingly patient waiting for her late night walks.

Thank you for your time and support of everyone mentioned, and all those I couldn't.

List of Abbreviations

GM	Geologos del Mundo
LS	Landslide
NLS	Non-landslide
IG	Information Gain
GR	Gain Ratio
CS	Chi-Squared
FS	Filtered Subset Evaluator

List of Figures

List of Tables

Abstract

The municipality of San Juan La Laguna, Guatemala is home to approximately 5,200 people and located on the western side of the Lake Atitlán caldera. Steep slopes surround all but the eastern side of San Juan. The Lake Atitlán watershed is susceptible to many natural hazards, but most predictable are the landslides that can occur annually with each rainy season, especially during high-intensity events. Hurricane Stan hit Guatemala in October 2005; the resulting flooding and landslides devastated the Atitlán region. Locations of landslide and non-landslide points were obtained from field observations and orthophotos taken following Hurricane Stan. This study used data from multiple attributes, at every landslide and non-landslide point, and applied different multivariate analyses to optimize a model for landslides prediction during high-intensity precipitation events like Hurricane Stan. The attributes considered in this study are: geology, geomorphology, distance to faults and streams, land use, slope, aspect, curvature, plan curvature, profile curvature and topographic wetness index. The attributes were pre-evaluated for their ability to predict landslides using four different attribute evaluators, all available in the open source data mining software Weka: filtered subset, information gain, gain ratio and chi-squared. Three multivariate algorithms (decision tree J48, logistic regression and BayesNet) were optimized for landslide prediction using different attributes. The following statistical parameters were used to evaluate model accuracy: precision, recall, F measure and area under the receiver operating characteristic (ROC) curve. The algorithm BayesNet yielded the most accurate model and was used to build a probability map of landslide initiation points. The probability map developed in this study was also compared to the results of a bivariate landslide susceptibility analysis conducted for the watershed, encompassing Lake Atitlán and San Juan. Landslides from Tropical Storm Agatha 2010 were used to independently validate this study's multivariate model and the bivariate model. The ultimate aim of this study is to share the methodology and results with municipal contacts from the author's time as a U.S. Peace Corps volunteer, to facilitate more effective future landslide hazard planning and mitigation.

1.0 Introduction and Objective

Lake Atitlán, an area in the western highlands of Guatemala, is home to approximately 200,000 people in twelve municipalities (Figure 1.1). Like much of Guatemala, natural hazards threaten this region throughout the year: earthquakes, volcanic activity, landslides, hurricanes/tropical storms and flooding. These hazards have the potential to both affect the people and the area, but landslides are far more common, as they can occur yearly with the seasonal rains.

Rain-induced landslides can be a small event, barely blocking the road, or they can occur by the thousands, with some volumes of material moved reaching 400,000 m^3 (Luna 2007). Landslide initiating rains usually result from hurricanes, tropical storms and tropical depressions. These high-yield precipitation events can bring precipitation in quantities that easily exceed the land's capacity to drain it safely, potentially causing countless deaths and damage to infrastructure. The most recent, damaging high-yield precipitation events that affected the Lake Atitlán region were Hurricane Stan in 2005 and Tropical Storm Agatha in 2010. One of the most devastating landslides on the lake occurred because of Hurricane Stan in the town Panabaj, killing 1400 (Luna 2007). In the local Mayan dialect "Panabaj" means "head of mud" (Luna 2007); a translation that indicated there is local understanding of the threat landslides have historically placed and continue to place on the residents of Lake Atitlán.

The author served as a Peace Corps Volunteer in one of municipalities on Lake Atitlán, San Juan La Laguna, from March 2010 to February 2012. Tropical Storm Agatha occurred at the end of May in 2010, allowing the author to experience firsthand the devastation such high-yield events can bring. During years of heavy precipitation, local people, adjust their travels to earlier in the day. They know that by the afternoon, after a few hours of steady rainfall, landslides will have occurred, hopefully only blocking roads or cutting through fields, but potentially burying people, cars, buses and infrastructure.

To help mitigate future impact of rain-induced landslides the organization Geológicos del Mundo (GM) quantified, as part of a larger project, the susceptibility of landslides for the Lake Atitlán watershed. GM translates as *World Geologists*. They evaluated landslide susceptibility using a database of landslides caused by Hurricane

Stan. Their study, which will be discussed later, developed a susceptibility map based on a bivariate statistical analysis and the influence of six physical attributes (GM 2012a).

The objective of this study is to develop a probabilistic landslide hazard assessment for the San Juan area using a simple, reproducible methodology. A multivariate approach is applied here because previous works (Suzen et al. 2004; Nandi et al. 2009) have shown it to be more successful than bivariate analysis and is more easily reproduced as new information becomes available. Different multivariate algorithms are evaluated to determine the optimized model for predicting landslide initiation. The results of this probabilistic map will be compared with the GM's (2012a) bivariate approach. Landslides from Tropical Storm Agatha in 2010 will be used to independently validate both statistical approaches. Lastly, the methodology and map of this study will be made available to the municipality.

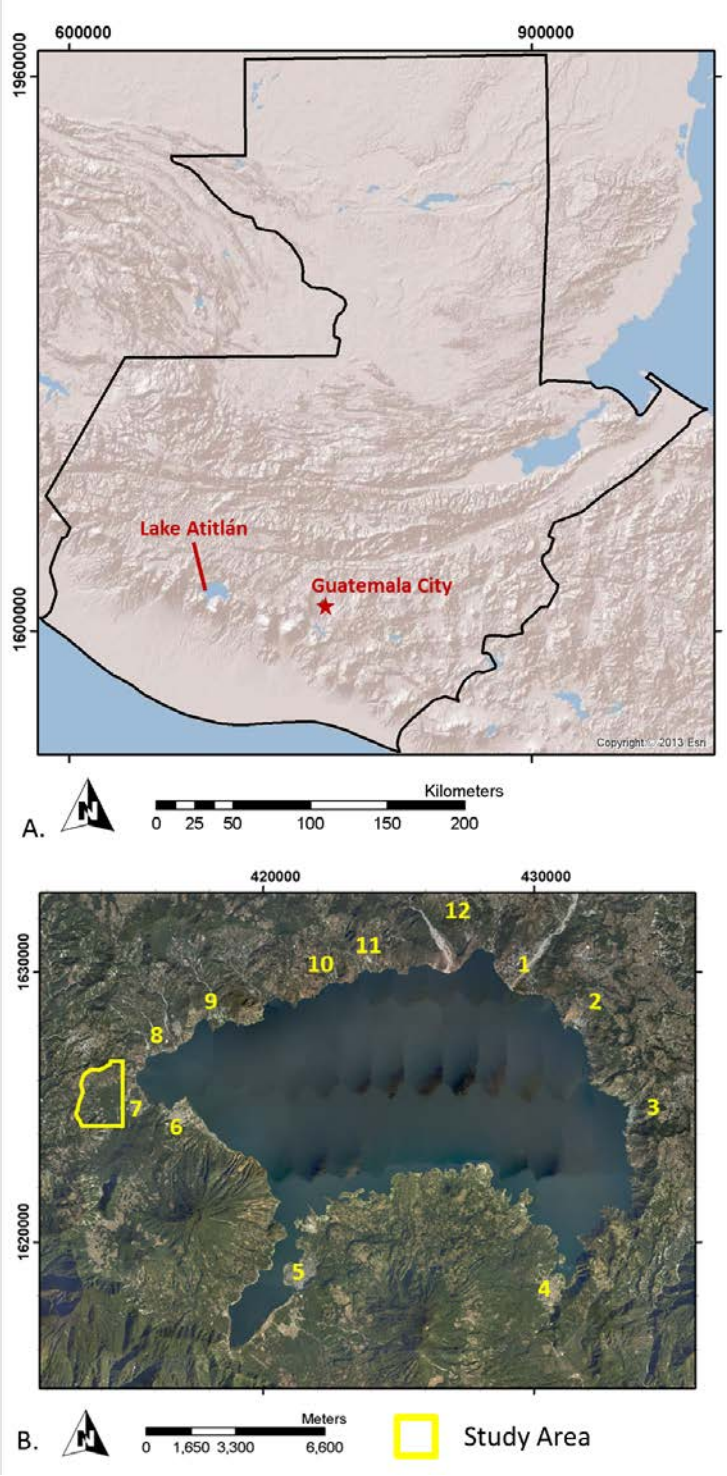


Figure 1.1: A.) A map of Guatemala, showing Lake Atitlán and where the study area is located, in the western part of the country. B.) The twelve major municipalities surrounding Lake Atitlán are numbered on the above orthophoto; 1) Panajachel, 2) Santa Catarina Polopó, 3) San Antonio Polopó, 4) San Lucas Tolíman, 5) Santiago Atitlán, 6) San Pedro La Laguna, 7) San Juan La Laguna, 8) San Pablo La Laguna, 9) San Marcos La Laguna, 10) Tzununá, 11) Santa Cruz La Laguna and 12) San Jorge La Laguna.

2.0 Previous Work

The municipal seat San Juan La Laguna sits on predominantly flat terrain, with a wide buffer of cultivated land between its infrastructure and the steep slopes prone to landslides during high-yield precipitation events. Silverman (2011) discussed how in other parts of Guatemala increases in population forced cultivation and building on steep slopes that had historically been avoided due to their susceptibility to landslides. This expansion that caused damages and death as a result of Tropical Storm Agatha in 2010, because there was no longer a buffer between the populace and landslide susceptible areas (Silverman 2011). Currently San Juan has enough room on flat ground to build, and while landslides definitely affect cultivated land, it is only the largest of landslides that tend not to happen yearly that affect infrastructure and lives. The town has been expanding, as noticed during the author's Peace Corps service. If this expansion continues unchecked, eventually the municipality will run into the same problems observed by Silverman (2011). It is integral that the slopes most at risk for landslide initiations be identified so the community can be better informed as to where they can expand with the most safety.

Silverman (2011) also discussed typical characteristics of different slope failures for the north-west portion of Guatemala. One such description is very similar to the slides found around San Juan, landslides that have a high length to width ratio and usually initiated at the top of ridges and flowed down slopes in drainage channels (Silverman 2011). While understanding the volume of potential material that can be mobilized in landslides, first it is informative to determine the areas most likely to fail. Once these susceptible areas are identified, the stream channels and topography can be used to model for flow. Different approaches have been used to identify landslide susceptible areas.

Numerous landslides susceptibility studies exist covering regions worldwide, employing varying statistical approaches. Van Westen et al. (1997) divides possible approaches for hazard zonation using GIS: heuristic qualitative approach, bivariate approach, multivariate statistical approach, and deterministic approach. Typically these approaches are applied to successively smaller study areas: 1:100,000 to 1:250,000, 1:25,000 to 1:500,000 and 1:2000 to1:10,000 (van Westen et al. 1997).

The heuristic qualitative approach has a low reproducibility from one study to the next because the hazard class is assigned based on the professional experience of the team (van Wester et al. 1997). This approach requires an expert understanding of different landslide mechanisms; the causative factors leading to slope failure. Landslides can be identified through fieldwork and/or aerial photo interpretation. The terrain conditions of all encountered landslides are evaluated, so, preliminary conclusions can be drawn about landslide causative factors in the given study area (van Westen et al. 1997). The potential combination of causative factors (slope angle, geomorphology, lithology, etc.) are determined by the professional experience (van Westen et al. 1997 and Aleotti et al. 1999). This preliminary analysis is then employed to establish weights for the different contributing factors (van Westen et al. 1997 and Aleotti et al. 1999).

The bivariate approach statistically evaluates the influence of one attribute or combination of attributes to the occurrence of landslides. Figure 2.1 shows the schematic of the bivariate approach (van Western et al. 1997). The combinations are chosen by the expert, making the bivariate approach part subjective and part objective (van Westen 1997). This method assumes that one attribute and distribution of landslides is not related to another attribute and the same distribution of landslides (van Westen 1997). For example, no correlation can be drawn between slope angle and geomorphological units. Depending on the attributes used, such an assumption is not problematic, but there are attributes that are dependent on each other. Van Westen et al. (1997) suggests the user needs to examine that data, making new attribute maps by incorporating the dependent ones together.

The multivariate statistical approach uses one of the two approaches: 1.) all available attributes for the study area are analyzed where they coincide with landslides by different regression techniques; 2.) the available attributes are crossed with distribution of landslide and non-landslides maps and a correlation is calculated via discriminate analysis (van Westen et al. 1997). Figure 2.2 is a schematic overview, detailing the multivariate technique. Multivariate statistical analysis is data driven, thus objectively classifying areas as to their potential for landslides (van Westen 1997). The only restriction of replicating a given multivariate study would be if the newly chosen study

site had the same available parameters, and it was assumed the new study area was susceptible to the same causative factors. Van Westen (2004) surmised that statistical approaches assume that each landslide, in a given area, experiences the same causative factors in the same way to cause slope failure, which could be elaborated upon to include landslides in different study areas.

Deterministic models evaluate slope stability by calculating a factor of safety for individual slopes. Where bivariate and multivariate methods can use solely data with a spatial extent, deterministic considers the depths and thicknesses of different material layers, and requires information on water table. Deterministic models further require that the failure mechanisms of each landslide, identified in the study area, be well understood (van Westen 1997).

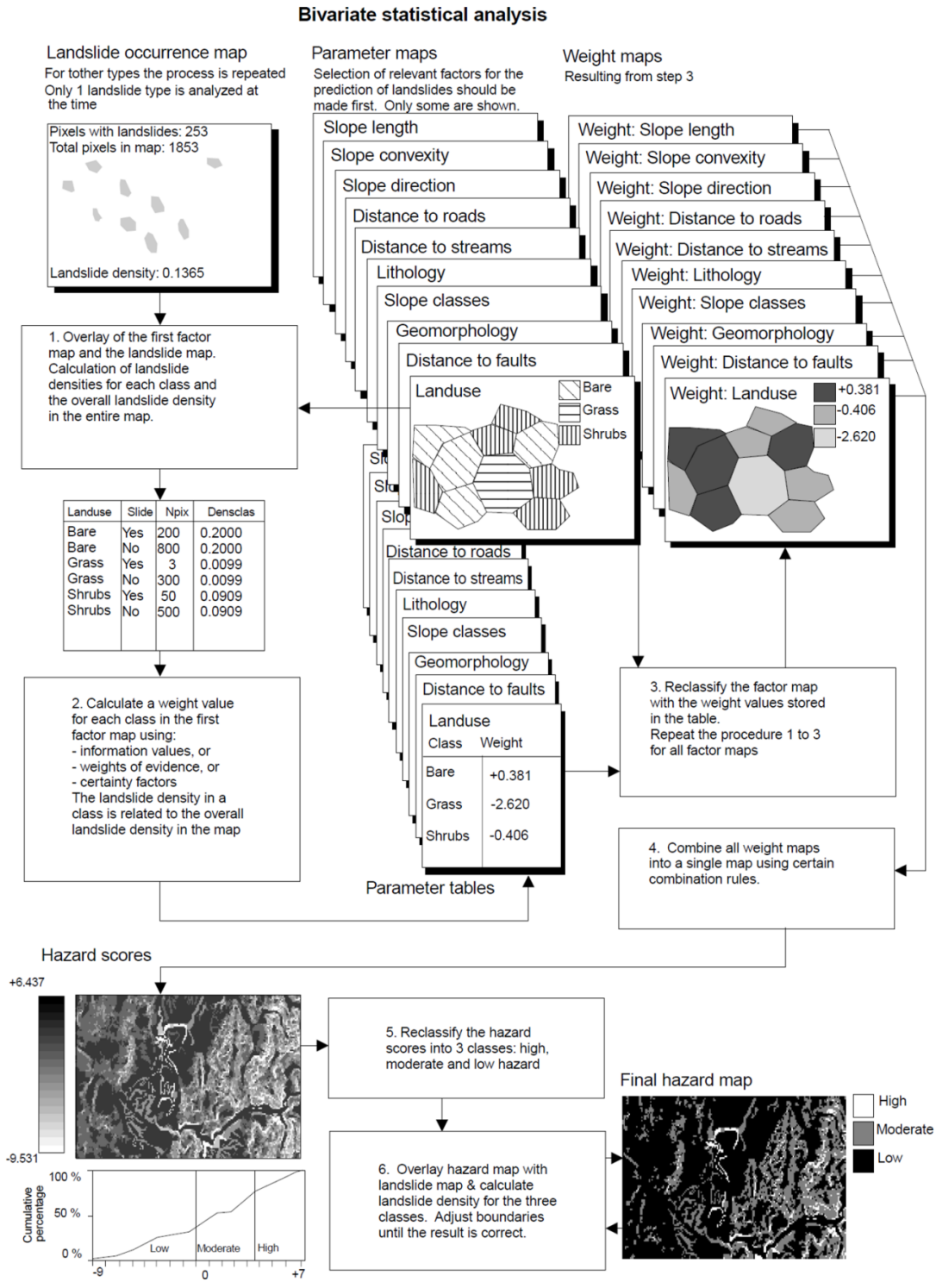
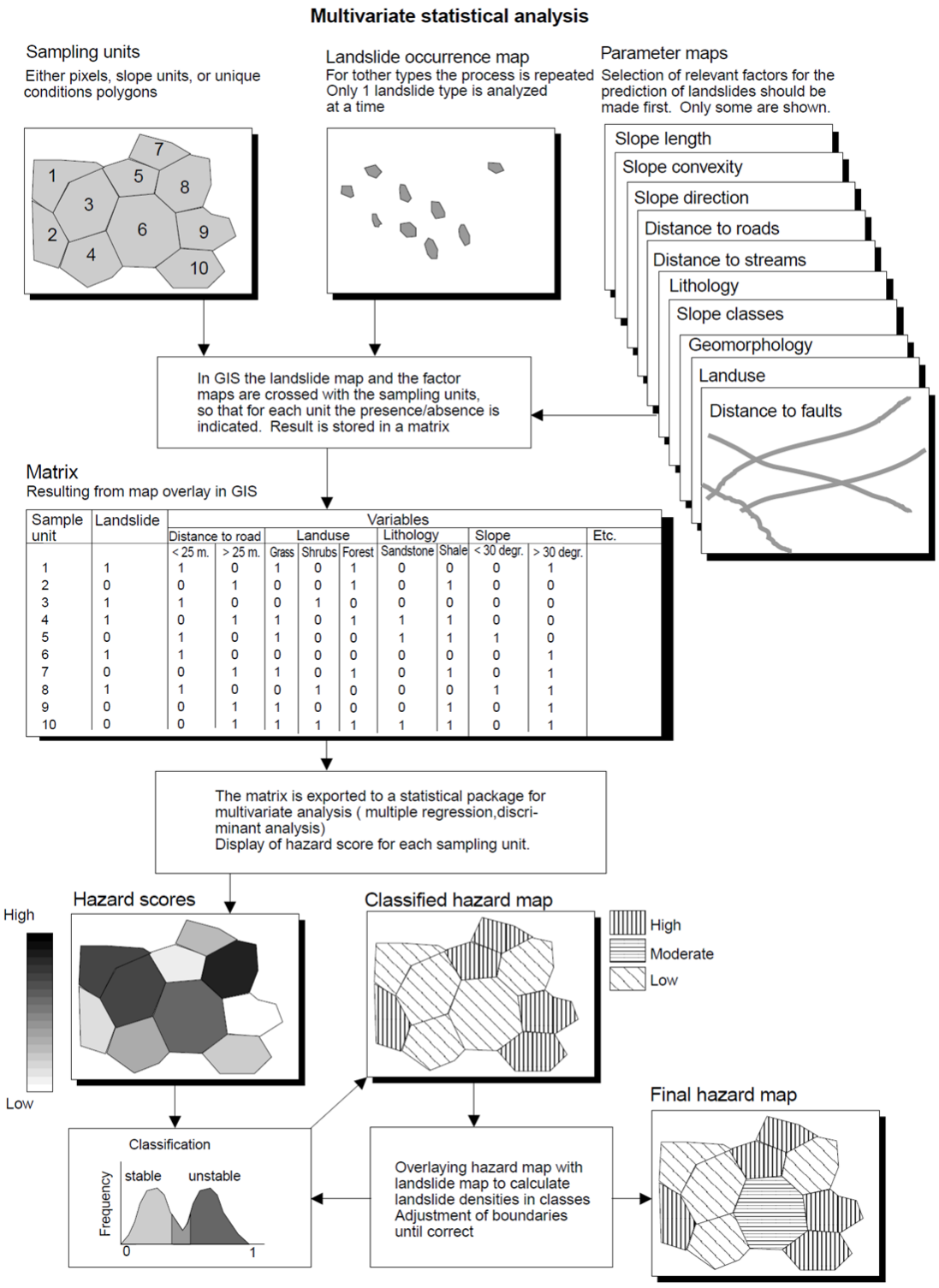


Figure 2.1: A schematic overflow of the bivariate statistical analysis approach using GIS. *Figure from van Westen et al. (1997).* Permission form is in appendix 3.



Sample unit	Landslide	Variables									
		Distance to road		Landuse			Lithology		Slope		Etc.
		< 25 m.	> 25 m.	Grass	Shrubs	Forest	Sandstone	Shale	< 30 degr.	> 30 degr.	
1	1	1	0	1	0	1	0	0	0	1	
2	0	0	1	0	0	1	0	1	0	0	
3	1	1	0	0	1	0	0	0	0	0	
4	1	0	1	1	0	1	1	1	0	0	
5	0	1	0	1	0	0	1	1	1	0	
6	1	1	0	0	0	0	0	0	0	1	
7	0	0	1	1	0	1	0	1	0	1	
8	1	1	0	0	1	0	0	0	1	1	
9	0	0	1	1	0	0	0	1	0	1	
10	0	0	1	1	1	1	1	1	0	1	

Figure 2.2: A schematic overflow of the multivariate statistical analysis approach using GIS. *Figure from van Westen et al. (1997).* Permission form is in appendix 3..

Of the four approaches described by van Westen (1997) only the bivariate and multivariate approaches are discussed further. A heuristic approach was not used here because it is heavily dependent on expert opinion, thereby not easily reproducible without an expert leading the investigation. Also, any recommendations made from such an approach could be potentially legally compromising, because they are based on the subjectivity of the expert(s) who conducted the study (van Westen 1997). The deterministic approach was not feasible for the study area due to certain attributes (thicknesses of materials covering the area, different strength properties of the same material and information on the water table) not being available. Deterministic approaches are also ideal for small study areas (van Westen 1997), and while the study area chosen here meets that requirement, another objective of this study is to develop a speedy and easily reproducible method. Deterministic analysis requires a landslide by landslide analysis, cumbersome when looking at a landslide database of one hundred plus. This study evaluates the potential of multivariate analysis to predict landslides in the region. The multivariate approach was chosen, in part, because in 2012 GM evaluated landslide susceptibility in the whole watershed of Lake Atitlán using a bivariate approach. Further, the multivariate methodology proposed in this study is faster; allowing for newer maps to be easily created as new and updated attribute information becomes available for the region. With a simpler method these maps could be created as local arises. Also, other studies show multivariate approach to be superior to the bivariate approach (Suzen et al. 2004; Nandi et al. 2009).

Though the Lake Atitlán region has experienced many devastating landslides, resulting from high-yield precipitation events, the only comprehensive landslide susceptibility study was completed in 2012 by GM (2012a). GM worked from June 2011 to March 2012 to develop landslide susceptibility maps for the Atitlán watershed (Figure 2.3). GM used a bivariate statistical approach to create the map because they had used a similar method in the metropolitan area of San Salvador in 2007 through 2008 (GM 2012a; Fernanadez-Lavado 2008).

GM looked at six attributes and weighed each individually to the distribution of the landslides in the watershed to determine how they influenced the occurrence of

landslides. For their study they considered the following attributes: geology, geomorphology, land use, slope angle, slope aspect and density of faults present. All the attributes were rasterized; for geology, geomorphology and land use they assigned each unit a unique number. For geology they used 1 through 32, for geomorphology 1 through 30 and for land use 1 through 19. While slope angle, aspect and density of faults were divided into five, four and nine ranges respectively. Each range was then assigned a unique number (GM 2012a). Table 2.1 shows the divisions GM used for the slope angle, aspect and density of faults.

Table 2.1: The unique number classifications for GM's study for the ranges of fault density (m/km^2), slope angle (degrees) and slope aspect (degrees).

Unique code	Fault Density (m/km^2)	Slope Angle (degrees)	Slope Aspect (degrees)
1	0 - 0.7	0 - 15	315 - 45
2	0.7 - 1.4	16 - 30	45 - 135
3	1.4 - 2.1	31 - 40	135 - 220
4	2.1 - 2.8	41 - 50	220 - 315
5	2.8 - 3.5	>51	
6	3.5 - 4.2		
7	4.2 - 4.9		
8	4.9 - 5.6		
9	5.6 - 6.3		

To establish the landslide distribution, GM used ArcGIS 9.3 and orthophotos taken by the Guatemalan government in 2006. GM identified 18,309 landslides of natural origin and 1,945 of anthropogenic origin (GM 2012a). Hurricane Stan occurred in October 2005, at the end of the rainy season, after which no significant storms affected the region causing no notable landslides. The orthophotos were acquired in January and February 2006 (Segeplan). GM assumed the 2006 orthophotos offered a reliable representation of the landslides caused by Hurricane Stan due to Stan marking the end of the high-yield events for the 2005 rainy season, and that the photos being taken before the rains began again in 2006 (GM 2012a).

GM calculated an index of susceptibility, 0 to 1, for each unique category of the six attributes (2012a). The first step to calculate this index required counting the number of landslides found in each unique category of the six attributes, and to determine the area that each category covered. Next, the ratio of number of landslides in an unique category to the area that category covered was found. Each attribute of *n* unique categories had one

category that had the highest ratio. To determine the index of susceptibility, this highest ratio was used to normalize all the other ratios in that attribute, by dividing all the ratios, of a given attribute, one by one, by the highest ratio calculated for that attribute. Thereby, the unique category that had the highest ratio of landslides to area and an index of susceptibility equal to 1 (GM 2012a).

They then summed all the attributes 5 different times, using different divisions of weight importance (Table 2.2). For each summation they determined natural breaks for 5 risk zones, with one being of lowest risk and five being of very high risk (GM 2012a)

Table 2.2: Five GM Hypotheses of Differently Weighted Attrubute Significance

	Weighted Importance of each Attribute (%)					
Hypothesis	Geology	Geomorphology	Slope	Aspect	Land use	Fault Density
1	15	15	50	10	5	5
2	15	10	60	5	5	5
3	10	10	70	3	2	5
4	10	5	75	3	2	5
5	10	6	74	3	3	4

GM determined the most accurate hypothesis, hypothesis 4, by calculating the number of landslides that fell in ranks 4 and 5, and the area that ranks 4 and 5 covered (2012a). Using hypothesis 4 they created their susceptibility map for the watershed (Figure 2.3). Their results will be compared to the results of this study.

While GM chose a bivariate analysis, many recent studies employ a multivariate approach. Chen and Wang (2007) employed a multivariate approach, using logistic regression to create a hazard map of Mackenzie Valley in Canada. They chose to evaluate the following attributes: bedrock, surface material, elevation, slope, aspect, dip angle, dip direction, distance to water system, and vegetation cover. They found the most important attributes for landslide hazard to be bedrock, surface materials, slope, and difference between surface aspect and dip direction of bedding rock (Chen et al. 2007). The exact combination of attributes identified as the most landslide causative depends on the study area; this study reviewed all the available data for the study area before choosing the ideal combination noted above.

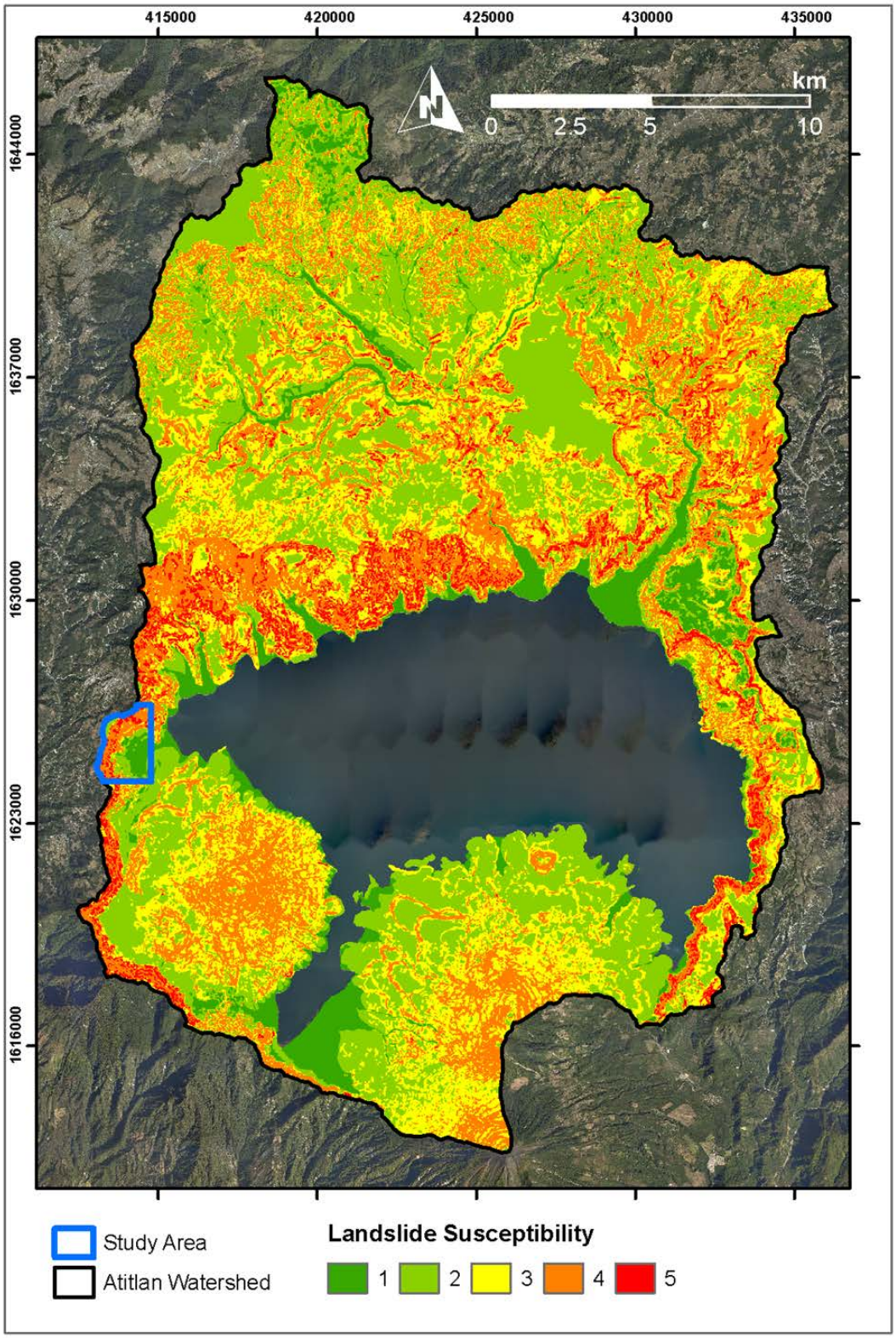


Figure 2.3: GM susceptibility map showing 5 hazard zones, (1) very low, (2) low, (3) medium, (4) high and (5) very high risk (GM 2012a).

Suzen et al. (2004) created susceptibility maps to compare the success of a bivariate versus multivariate analysis, using logistic regression. They evaluated different combinations of thirteen attributes: lithology, distance to fault, fault density, elevation, distance to drainage line, drainage density, distance to ridges, slope angle, slope aspect, distance to settlement, distance to power lines and roads, distance to local major highway and land cover. Suzen et al compared the two susceptibility maps quantitatively by the susceptibility classes and corresponding densities of landslides, and comparing the susceptibility classes to their spatial delineations. Suzen et al. (2004) determined the multivariate approach to be more reliable.

Nandi et al. (2009) also compared a bivariate approach to a multivariate approach, and like Suzen et al. (2004) used logistic regression, in the Cuyahoga River watershed of northeastern Ohio. They evaluated the following attributes: slope angle, soil type, soil erodibility, soil liquidity index, land-cover pattern, precipitation, and distance to stream (Nandi et al. 2009). To evaluate the success of both susceptibility maps Nandi et al. calculated area under the curve (AUC), and found the AUC of the bivariate analysis to be 0.59 and the multivariate analysis to be 0.81. Yesilnacar et al. (2005) describes AUC as the ability of the model to correctly identify whether the event, in this case landslides, occurs or not, based on some number of attributes, thereby matching or disagreeing with the known reality of the event, landslide or non-landslide. Again, the multivariate technique of logistic regression was found to be superior (Nandi et al. 2009).

Yesilnacar et al. (2005) conducted a comparative analysis, comparing logistic regression to neural networks for landslide susceptibility mapping in Turkey. They completed a forward stepwise logistic regression using a chi-square test to determine important attributes (Yesilnacar et al. 2005). The logistic regression used 14 attributes: aspect, distance to drainage, distance to ridges, distance to road, drainage density, elevation, fault density, geology, land cover, road density, slope, stream power, topographic wetness index and subwatershed basins. The neural network incorporated the same 14 attributes plus distance to fault planes, plan curvature, profile curvature, slope length and surface area ratio (Yesilnacar et al. 2005). The AUC of the logistic regression was 0.76 and 0.89 for the neural network (Yesilnacar et al. 2005).

Another possible multivariate approach is Bayesian networks. Ozdemir (2011) used a Bayesian network in the Sultan Mountains of Turkey. Ozdemir (2011) had 19 attributes that could be potentially landslide causing: geology, relative permeability, land use/land cover, precipitation, elevation, slope, aspect, total curvature, plan curvature, profile curvature, wetness index, stream power index, sediment transport capacity index, attitude, distance to drainage, distance to fault, drainage density, fault density and spring density map. Using chi-square statistics 4 different models were created from different attribute combinations, determining the accuracy of each by using AUC. Ozdemir (2011) determined the combination of geology, profile curvature, slope and spring density to be the most accurate for the study area with an AUC of 0.92.

Pradhan et al. (2010) used the Bayesian probability model of weights-of-evidence. They determined the ideal attribute combination to be plan curvature, distance from drainage, distance from lineament, geology and land cover, based on the highest AUC obtained in the modeling, 0.80.

Saito et al. (2009) used J48 decision tree, a reimplementation of the C4.5, was used with the following attributes: geology, elevation, slope, profile curvature, plan curvature, dissection height and undissection height. To evaluate accuracy Saito et al. employed the kappa coefficient (k), where 1 is indicative of perfect agreement between model and known database of landslides (Cohen 1960; Hoehler 2000). The kappa coefficient of this study was found to be 0.61, at the 99% confidence level (Saito et al. 2009).

Suzen et al. (2004) and Nandi et al. (2009) found multivariate logistic regression to be superior to a bivariate statistical analysis. They used landslide density per hazard class area and AUC respectively. Cheng and Wang (2007), Pradhan et al. (2010) and Ozdemir (2011) determined the ideal attribute combination for their model in their study area. Some of the attributes were the same between the studies. Slope and geology appeared in all three. Pradhan et al. (2010) and Ozedmir (2011) used a derivative of curvature, plan curvature and profile curvature respectively. Even though these three studies found different ideal combinations of landslide causative attributes, that follows van Westen's (2004) observation that not all landslides, even in the same region,

experience the same factors the same way. This study will optimize for the ideal combination of landslide causative attributes and identify the ideal algorithm for predicting landslide areas.

3.0 The Study Area

The watershed of Lake Atitlán is found between longitudes 93° 48'W and 93° 35'W, latitudes 14° 53'N and 14°, covering 546 km^2, with the lake accounting for 123 km^2 (Figure 3.1). The maximum depth of lake is 324 m (CONAP 2007). Steep slopes form the walls of this watershed, which were created by three caldera building events as seen in Figure 4.4 (Newhall 1987a). The southern boundary of the lake is marked by three strato-volcanoes, Atitlán, Tolimán and San Pedro (Figure 3.1). The elevation of the watershed ranges from approximately 1,500 meters at the lake edge to just over 3,500 meters at the summit of volcano Atitlán. Over two thousand drainage channels, of varying size (Figure 3.1), feed the lake.

The perimeters of the study area are drawn in yellow in Figure 3.1, overlaying orthophotos taken by the Guatemalan government in early 2006. The study area covers 4.05 km^2. The area excludes much of the town infrastructure of San Juan La Laguna, which is to the east and south east of the perimeter. The study area was chosen to focus on the steep slopes surrounding the town, where landslides originate almost yearly (Figure 3.2). The town of San Juan was built on relatively flat terrain, while the surrounding slopes creep to above 60°.

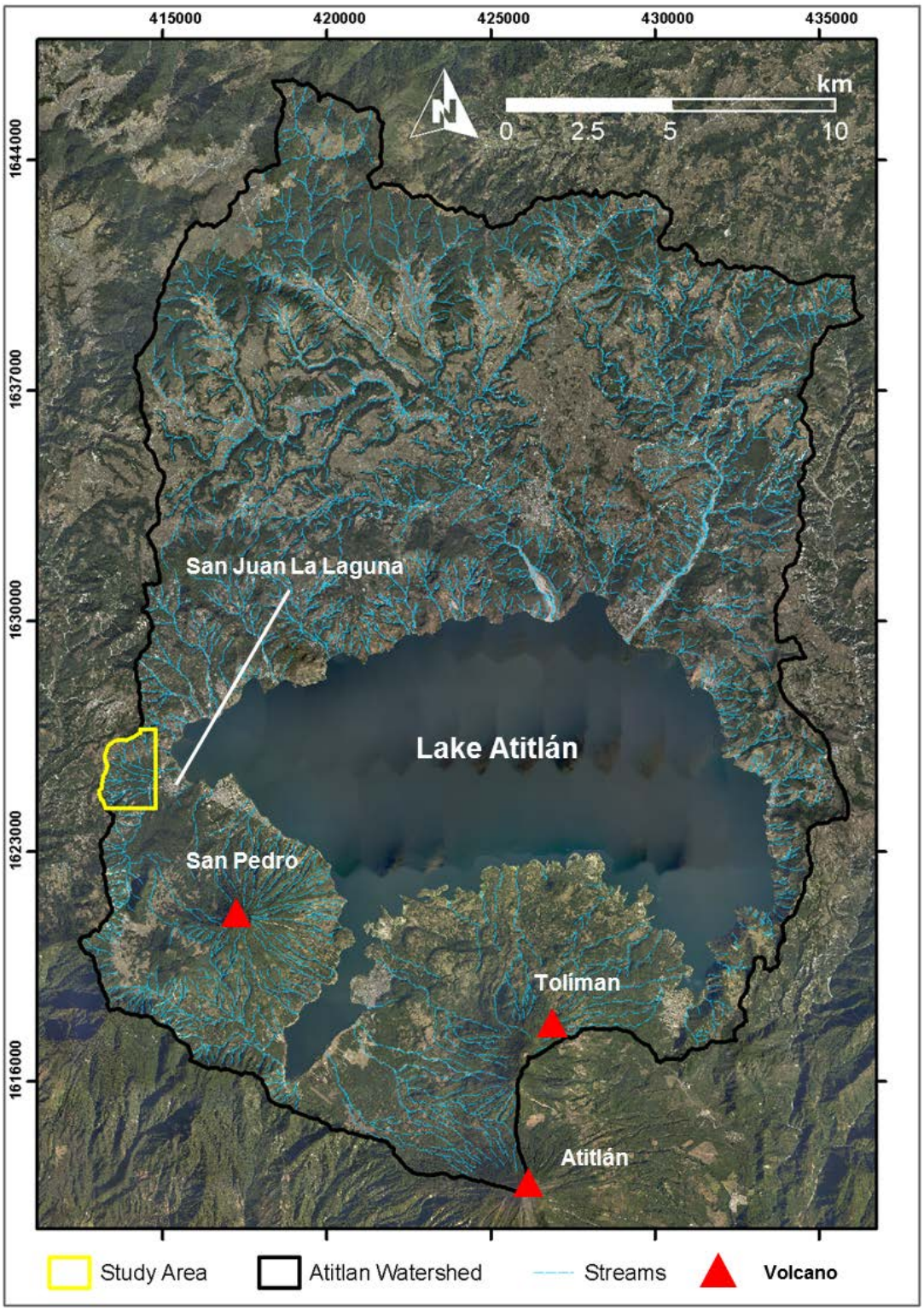


Figure 3.1: Map of Lake Atitlán watershed and study area. The watershed is outlined in black and the study area in yellow. Three volcanoes, San Pedro, Tolíman and Atitlán, sit on the southern shore of the lake. The drainahe network is shown in blue.

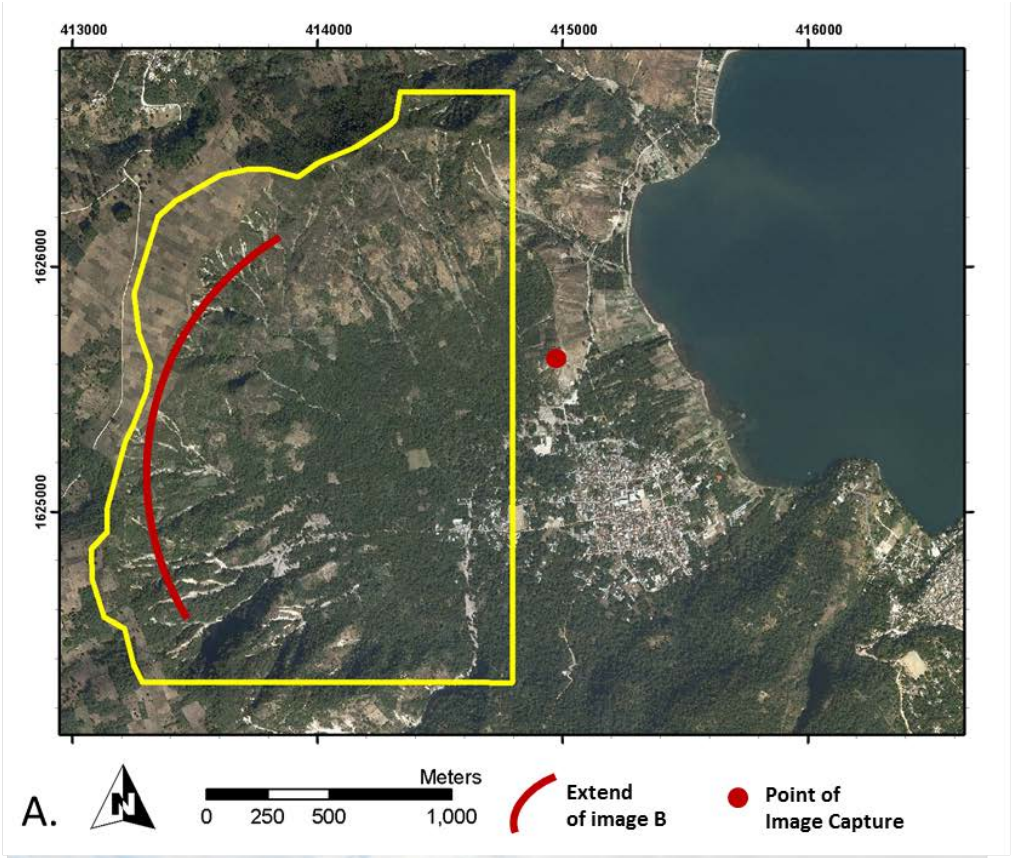


Figure 3.2: A) The yellow rectangle indicates the extent of the study area, which includes a red arc of the approximate extent of Figure 3.2b. The red dot is the approximate location where the image was captured. B) Looking west, these are the steep slopes that flank the town. The landslides visible here resulted from Storm Agatha in 2010. *Photo taken by author.*

4.0 Developing the Spatial Database

To perform the multivariate analysis a data set of landslides and non-landslides points had to be created, as well as, identifying all available potential landslide causative attributes. A large landslide inventory, resulting from Hurricane Stan in 2005, for the watershed was made available to the author by Laura Nunéz Álvarez of GM. The landslides from this database that fell within the study area were used as a spatial reference in the final choosing of landslide points (LS points) for the modeling (Figure 4.1). Non-landslides points (NLS points) were randomly chosen within the bounds of the study area. GM also provided shapefiles of the following primary attributes: geology, geomorphology, drainage channels and faults (2012a). A fifth primary attribute, land use, was available through the Guatemalan Government (Segeplan 2006). Shortly after Hurricane Stan, the government of Guatemala took orthophotos of the whole country, at a spatial resolution of 0.5m, and created 20m contour lines (2006). The Guatemala Government also made the digital elevation model of the whole country available, from which the following secondary attributes were derived: slope angle, slope aspect, curvature, plan curvature, profile curvature and topographic wetness index (TWI).

4.1 Inventory of Landslide and Non-Landslide Points

The GM landslide inventory for the watershed consisted of 18,309 landslide points (Figure 4.1). GM developed the inventory using the 2006 orthophotos, taken after Hurricane Stan of October 2005, and confirmed, as needed, with fieldwork and Google Earth images (MAGA 2006; GM 2011). Each landslide in the database was marked towards the top of the scarp, roughly in the middle. These points do not necessarily represent the initiation point of slope failure, but rather just a consistent spot to mark from slide to slide. Rather than build the model on arbitrary centroids at the top of the landslide scarp, the ArcTool, *Spatially Balanced Points*, was used to help randomly vary the placement of landslide points (LS points) along the scarp.

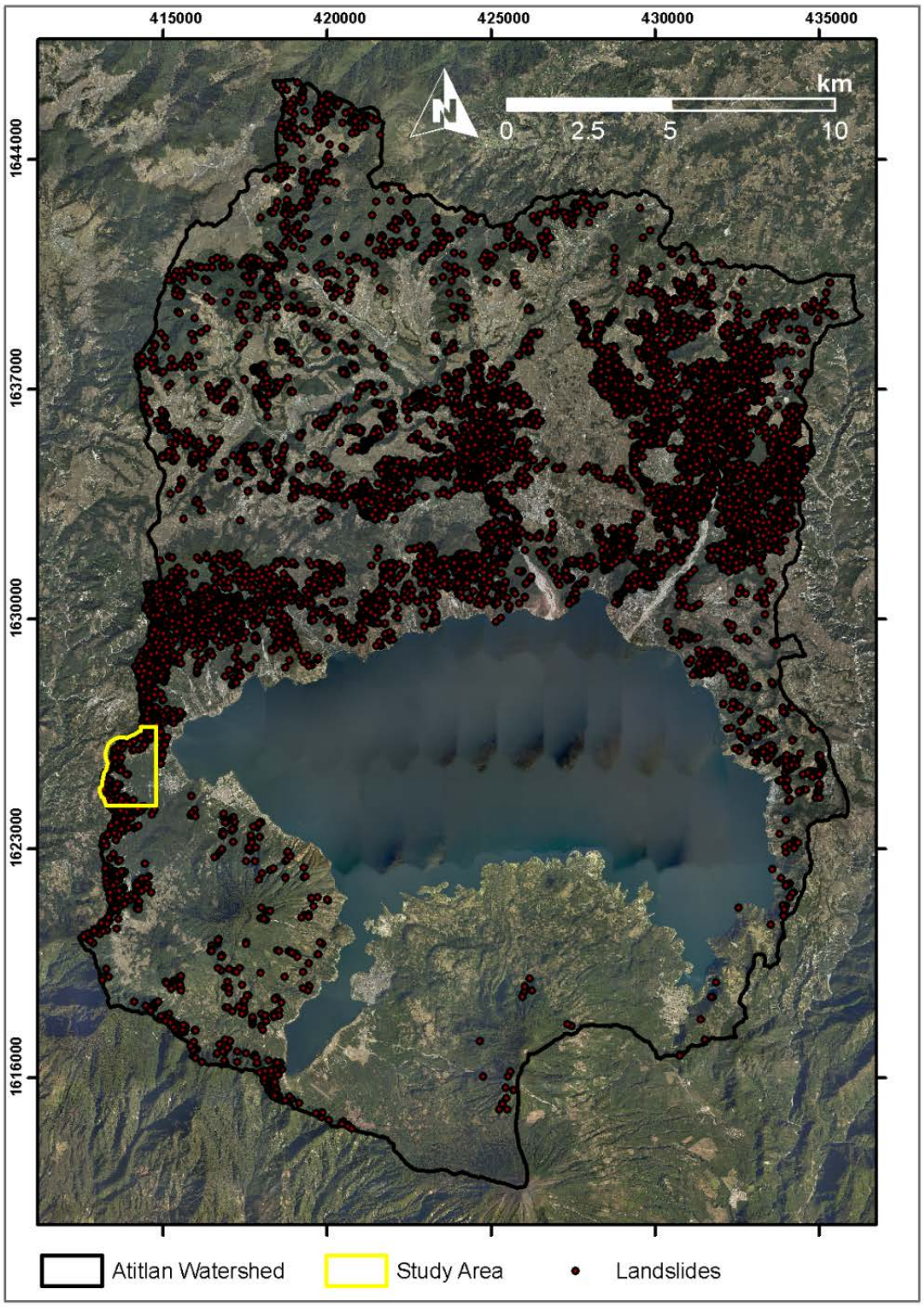


Figure 4.1: The 2006 orthophotos showing the watershed of Lake Atitlán, outlined in black, and all the landslides documented to have occurred as a result of Hurricane Stan. Note also the yellow square delimitating the study are discussed in this research.

The tool *Create Spatially Balanced Points* can create a table of geographically referenced points based on specified spatial limitations, using the Reverse Randomized Quadrant-Recursive Raster algorithm (ESRI 2012a). The ArcGIS tool was developed using the methods proposed by Theobald (2007) and Stevens et al. (2004). The tool selects points based on different spatial criteria. In choosing the LS points, the spatial options were constrained to a 5 meter radius around the centroids GM identified. Five meters was chosen because that approximated the average width of the landslides that occurred in the study area. The tool was directed to choose 200 points (Figure 4.2).

For non-landslide points (NLS points) the same ArcTool was used. The tool was instructed to select 1,000 points, with no two points closer than 5 meters, again due to the average width of the landslide in the study area. Of the 1,000 points, 650 intersected all the available probable landslide causative attributes (Figure 4.2).

The ArcTool *Create Spatially Balanced Points* chooses within the limitations of the spatial extent specified. The tool cannot identify if the pixels it is choosing fall on a LS or NLS point. When choosing the LS points within a 5 meter radius of the GM identified centroids, some of those radii did not fall completely on landslide pixels, meaning some of the 200 LS points were in fact NLS. Conversely, some of the 650 NLS points were in fact LS points. Keyport et al. (2013) discussed the possibility of remotely determining areas of landslide and non-landslide using only the spectral properties of the orthophoto. The extent of the study area overlapped the portion evaluated in this study.

To correct for misclassified LS and NLS points all the points were overlain on the orthophotos. Every LS point that was misclassified was either moved slightly or deleted, depending on if there were a cluster of pixels already on that landslide. Every NLS point that fell on a LS pixel was deleted. Data set A of 200 LS and 650 NLS points became Data set B of 191 LS and 560 NLS (Figure 4.2 and Table 4.1). Figure 4.3 shows the shift of the LS and NLS points from Data Set A to Data Set B.

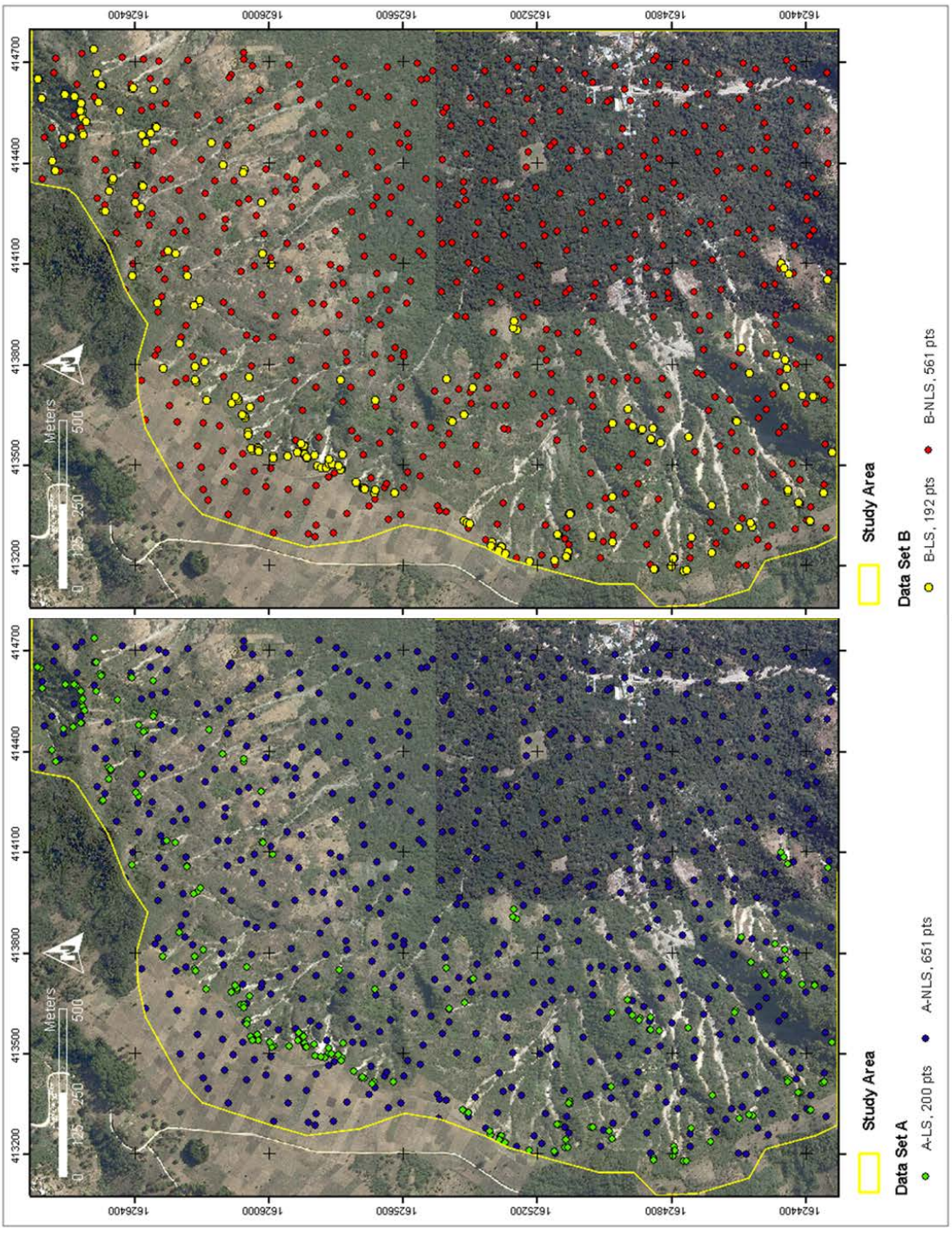


Figure 4.2: a.) Data set A, showing the 650 NLS points in blue and the 200 LS points in green on the MAGA orthopohots. b.) Data set B, displaying the filtered 561 NLS and 190 LS points on the MAGA orthophotos.

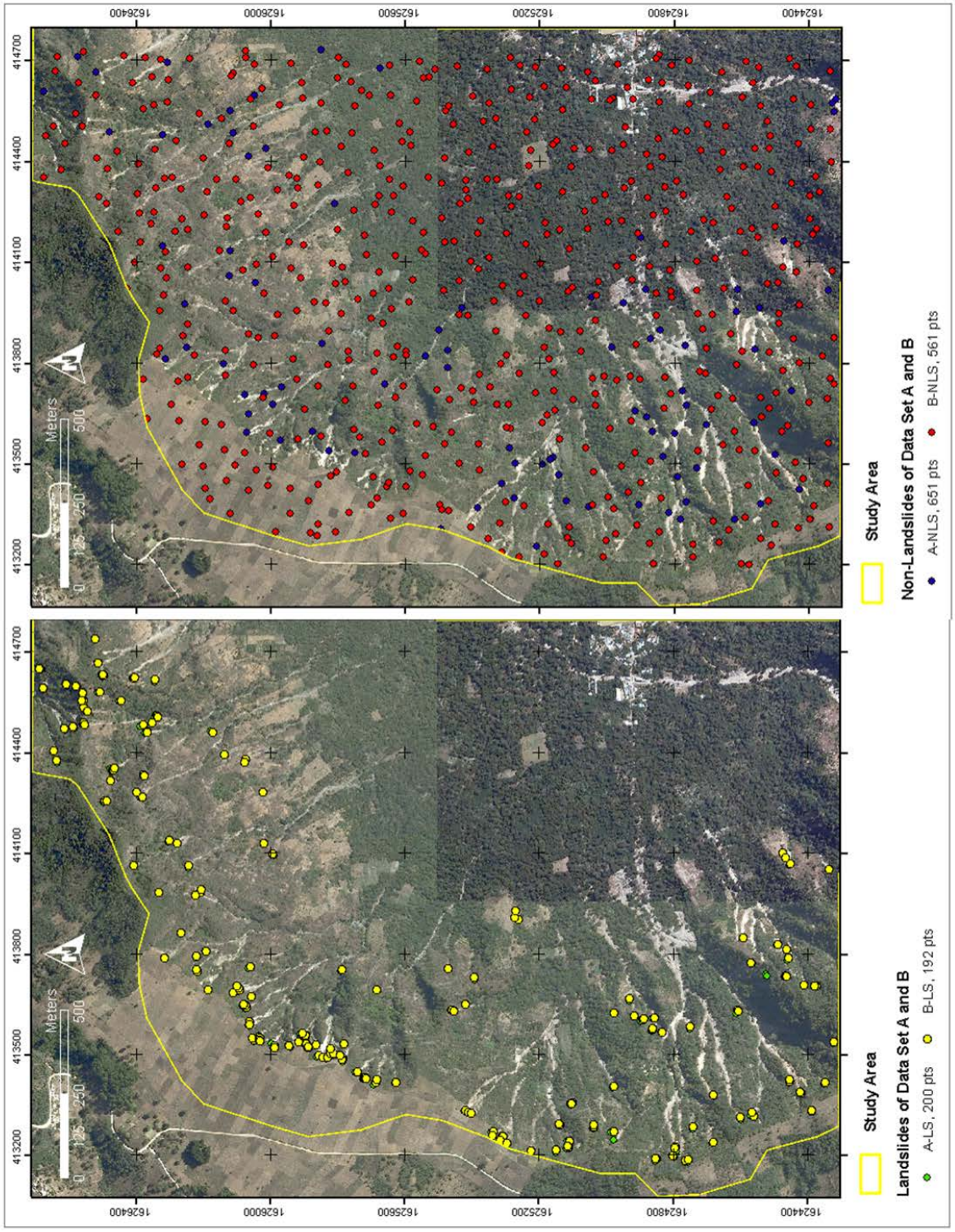


Figure 4.3: a.) Comparing the NLS points from data set a (A-NLS) and data set b (B-NLS). When only green points are visible that is an indication they were deleted in data set B. b.) Comparing the LS points from data set a (A-LS) and data set b (B-LS). When only blue points are visible that is an indication they were deleted in data set B.

Table 4.1: Data Set A includes NLS and LS points that were chosen with the ArcTool Spatially Balanced Points in ArcGIS. Data set B resulted from filtering data set A, by either eliminating or moving the NLS and LS points to truly correspond with their classifications

Data Set A	
200 LS points	650 NLS points
Data Set B	
191 LS points	560 NLS points

4.2 Identifying Primary and Secondary Attributes

Picking the correct combination of landslide causative attributes usually has more to do with the data available for the study area than the actual ideal combination that would best predict a LS point if all imaginable attributes were available. Van Westen et al. (2008) described why some attributes may be landslide causative factors (Table 4.2).

Table 4.2: A summary of potential properties that might relate why these attributes could be landslide causative. Table modified from van Westen et. al (2008).

Attribute Name	**Landside Causative Properties**
Geology	Lithological classifications that can incorporate engineering characteristics
Geomorphology	Subdivides the terrain into zones
Distance to Faults	Typically found indicative of landslide in predictive mapping
Distance to Streams	First order streams are buffered
Land use	How the land is utilized influences the probability of landslide
Slope Angle	Important factor when gravitational movements are in question
Slope Aspect	Could reflect differences in soil moisture and how the slope is vegetated

Van Westen et al. (2008) does not include curvature, profile curvature or plan-form curvature, which are included in this study. Nor is TWI included, but van Westen et al. (2008) does touch on other hydrological parameters that attempt to account for rain fall.

GM (2012a) chose to use the attributes geology, geomorphology, distance to streams, fault density, slope angle and slope aspect because the same combination was used in San Salvador, El Salvador for the same bivariate approach (Fernanadez-Lavado 2008). Of the other studies discussed earlier (Suzen et al. 2004; Yesilnacar et al. 2005; Chen et al. 2007; Nandi et al. 2009; Saito et al. 2009; Pradhan et al. 2010; Ozdemir 2011), they all used a combination of different attributes, depending on what was available (Table 4.3). Some of the studies concluded which attributes proved to be the most crucial in predicting LS and NLS, however, not all the studies excluded attributes selectively to identify those with the most influence.

Table 4.3: Commonly used attributes in other studies. Some studies included more attributes than listed here, but they were omitted from the table because those attribtues were not available for this study area.

	This Study	GM	Suzen	Yesilnacar	Chen	Nandi	Saito	Pradhan	Ozdemir
Geology	**x**	x	x	x	x	x	x	x	x
GeoMorph	**x**	x			x				
Land Use	**x**		x	x		x		x	x
Dist to Streams	**x**	x	x	x	x	x		x	x
Dist to Faults	**x**		x	x				x	x
Landuse	**x**				x				x
Slope	**x**	x	x	x	x	x	x	x	x
Aspect	**x**	x	x	x				x	x
Curvature	**x**								x
Plan Curv.	**x**			x			x	x	x
Prof. Curv.	**x**			x			x		x
TWI	**x**			x					x

Selecting attributes to be included in a study solely because other studies found certain attributed to be important can be misleading. Van Westen (2004) cautioned that statistical approaches assume that each landslide in a given area experiences the same causative attributes in the same way to cause slope failure. This can be taken as a cautionary step against further assuming the combination of attributes that proved landslide causative in one area automatically translates to another geographical location. All the available data attributes for the study area were gathered and first pre-processed with multiple attribute evaluators available in the open source data mining software Weka (Hall et al. 2009). The pre-processing looked at a given attribute and ranked that attribute by how well the different attribute categories were able to group the LS and NLS points separately. After all the attributes were pre-processed they were then ranked by how well the attributed grouped LS and NLS points. The attributes that had larger stratification of the two point types were determined to have a higher rank over the other attributes. The pre-processing was used to inform the attributes selected and eliminated to create the ideal model for predicting landslide initiations. Different attribute evaluators proved more accurate in the attribute combination it determined to have the most landslide causative influence.

4.3 Geology

Lake Atitlán is at the bottom of a caldera, rimmed by steeply sloped walls and flanked to the south and southwest by three andesitic strato-volcanoes: San Pedro, Tolíman and Atitlán (Figure 4.4) (Newhall 1987a). The caldera, that contains the lake, results from three cycles that formed and filled three calderas, Atitlán I, Atitlán II and Atitlán III (Figure 4.4). Each cycle experienced strato-volcano development and high-volume silicic eruptions (Newhall 1987a).

The three cycles occurred during the following time periods respectively, 14 – 11 mya, 10 – 8 mya and 1 – 0 mya. Intrusions of mafic and intermediate lava, and eruptions of these lavas from strato-volcanoes marked the beginning of each cycle (Newhall 1987a). Eventually low-volume silicic plutons developed, part of which would erupt as ash flows (Newhall 1987a). Over a couple million years, near the surface, a larger body of silicic melt would develop, ultimately erupting as high-volume ash flows, creating the caldera corresponding with the cycle (Newhall 1987a). Each caldera would then be filled in by material ejected during that cycle and early material being erupted as the following cycle would begin (Newhall 1987a).

Christopher Newhall discussed the geologic history of the region in more detail in Newhall (1987a) and Newhall et al. (1987b), which GM updated in 2011 by developing four 1:25,000 geologic maps that covered the watershed of Lake Atitlán (GDM 2012a). The watershed was broken up into four quadrants, with the study area falling into the southwestern portion, entitled Santiago Atitlán. The geological map of the watershed is included in appendix 1.

The geologic map, entitled Santiago Atitlán, includes one of the three quaternary strato-volcanoes, San Pedro and portions of the western flanks of the other two volcanoes (GDM 2012a; GM 2012b). The southeastern quadrant, San Lucas Tolíman, covers one of the volcano Tolimán completely and includes portions of the volcano Atitlán's northwest flank. The majority of volcano Atitlán falls outside the watershed. The watershed bisects the saddle between volcanoes Tolíman and Atitlán, including only the northwestern part of the Atitlán volcano (Figure 3.1).

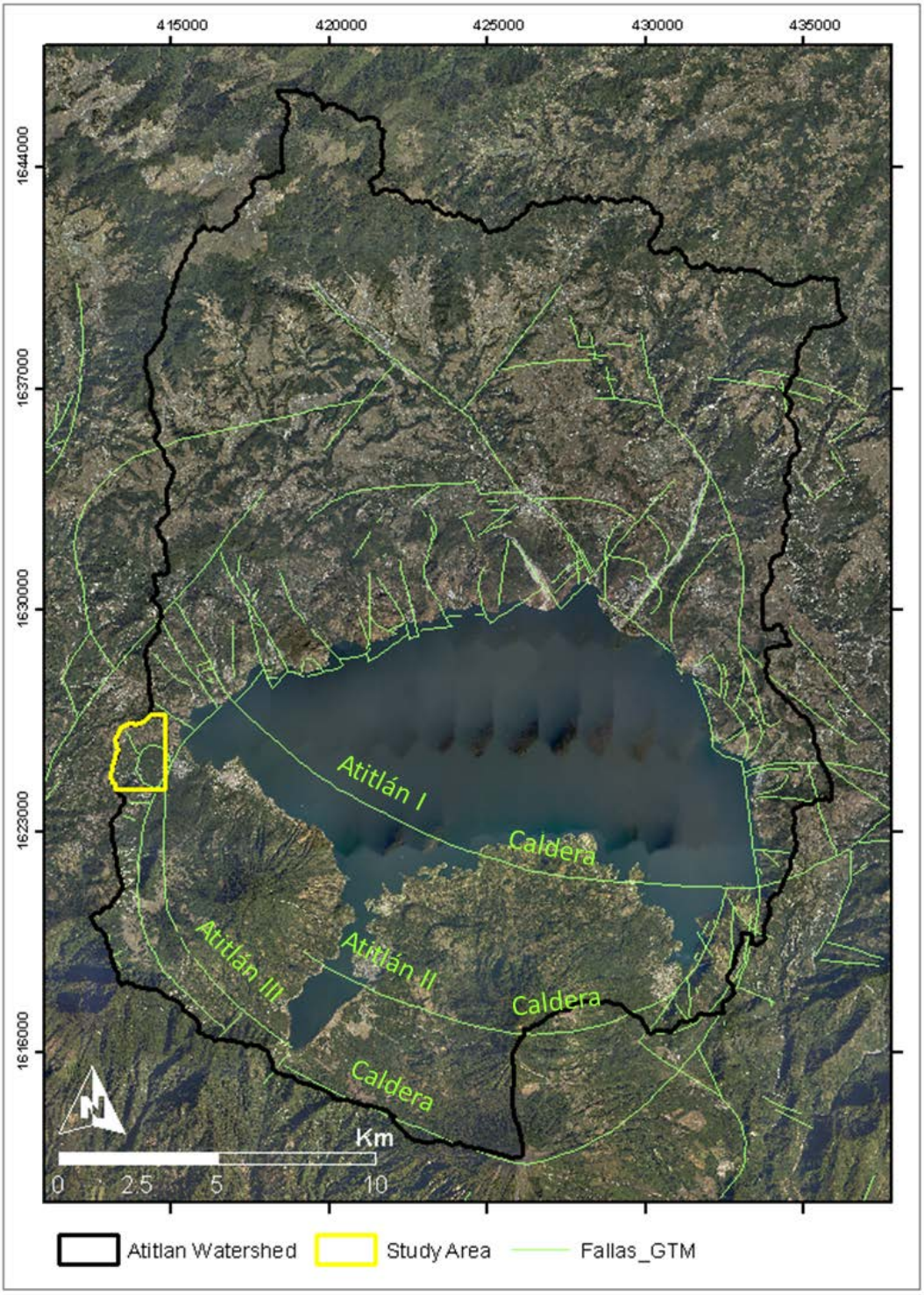


Figure 4.4: Map of Lake Atitlán showing the 3 calderas, Atitlán I, II and III, that exist in the region and the three stratovolcanoes present on the southern side of the lake today. The red rectangle indicates the extent of the study area. *Caldera Shape file source: GM 2012a.*

The three volcanoes began developing after the caldera Atitlán III was created, with San Pedro being the first to grow, followed by Tolíman and Atitlán respectively (Newhall 1987a). The most recent volcanic activity occurred on the Atitlán volcano with reports beginning in 1826 and ending in 1956 (Newhall 1987a).

The study area includes the following 7 lithologies of the 32 found in the Santiago Atitlán quadrant (appendix 1):

Table 4.4: Each lithological unit in the study area is described.

Lithological Unit	Description
Qa	Lava, mud and andesitic pyroclastic flows from quaternary complex stratovolcanoes.
Qal1	Sand, silt, clay, gravel, cobbles and boulders (alluvial fans).
Qcol2	Boulders, cobbles, gravel, sand, silt and reworked pumice and clay (colluvium).
Qpa4	Pumice, pyroclastic deposits post the Los Chocoyos event
Qpf3	Pumice, pyroclastic flows from the Los Chocoyos event
QTa	Andesitic lava, mud, and pyroclastic flows associated with tertiary-quaternary stratovolcanic complexes.
Tg	Granite and Granodiorite

Unit Qta, consisting of lava, mud and pyroclastic flows, covers the largest area, and makes up the portion of the study area that has the steepest slopes (Figure 4.5). Other steep slopes are found in the area of Qpa4, also a deposit comprised of pycroclastic material. Qcol2 is found on the gentlest slops in the study area (Figure 4.5).

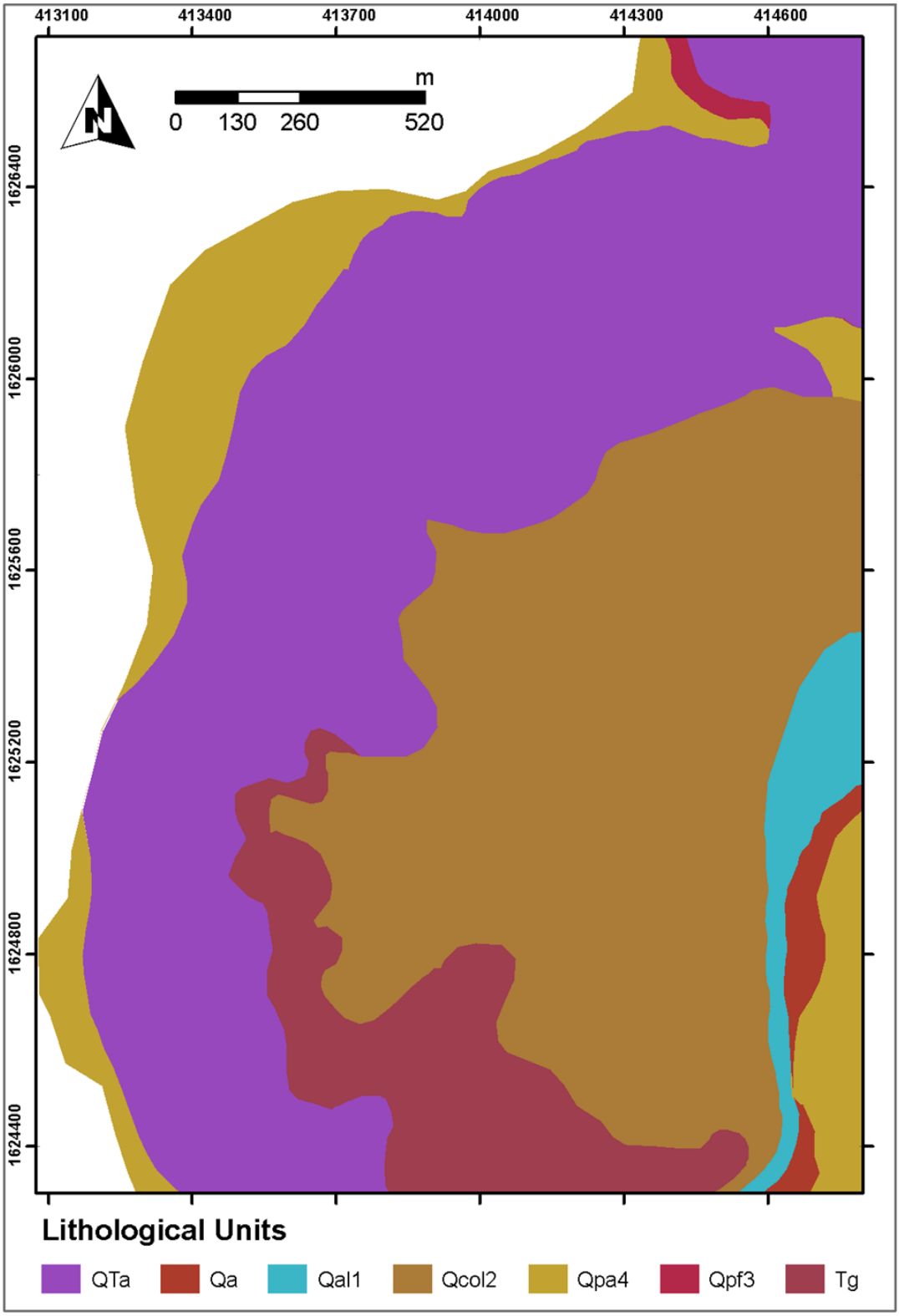


Figure 4.5: The geological units found in the study area.

4.4 Geomorphology

GM also developed geomorphological maps for the region (GM 2012). The geomorphological map of Santiago Atitlán has 30 units (appendix 1). The study area only includes the following 8 geomorphological units:

Table 4.5: Descriptions of the 8 geomorphological units found in the study area.

Lithological Unit	Description
AA	Alluvial Fan
B	Ravines
CLD	Degraded Lava Flows
FC	Lava Flow Front
LBCIII	Caldera Rim
Pm	Piedmont
SAP	Pumice/Ignimbrite Plateau
SAPD	Non-indurated Pumice/Ignimbrite Plateau

Unit LBCIII, weathered lava flows, covers the largest area, and makes up the portion of the study area with the steepest slopes (Figure 4.6). Piedmont, unit PM, resides on the gentlest slopes of the study area (Figure 4.6).

4.5 Distance to Stream

The study area contains 19 tributaries that were mapped by GM on their geological map (Figure 4.7). Most of the tributaries eventually merge together, as can be observed in the western part of the study area in Figure 4.7. Buffers were created, using ArcTool *Buffer*, around each of the rivers in the following increments in meters, 0-25, 25-50, 50-75, 75-100, 100-150, 150-200, 200-300 and 300-500. The ranges for the buffers were chosen in part because these ranges were typical in similar studies and due to how close the drainage channels were to each other (Chen 2007; Pourghasemi 2012).

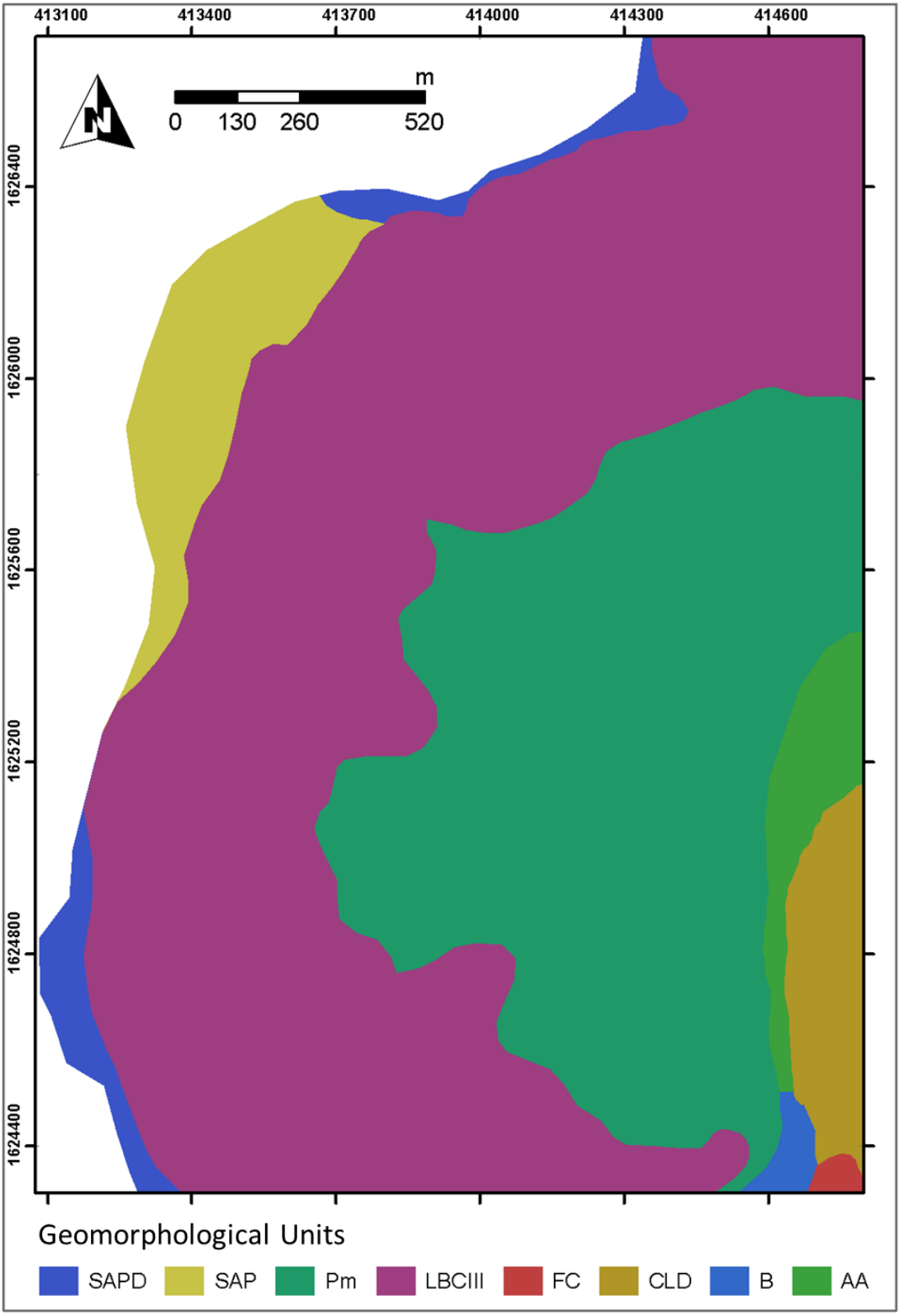


Figure 4.6: The geomorphological units included in the study area.

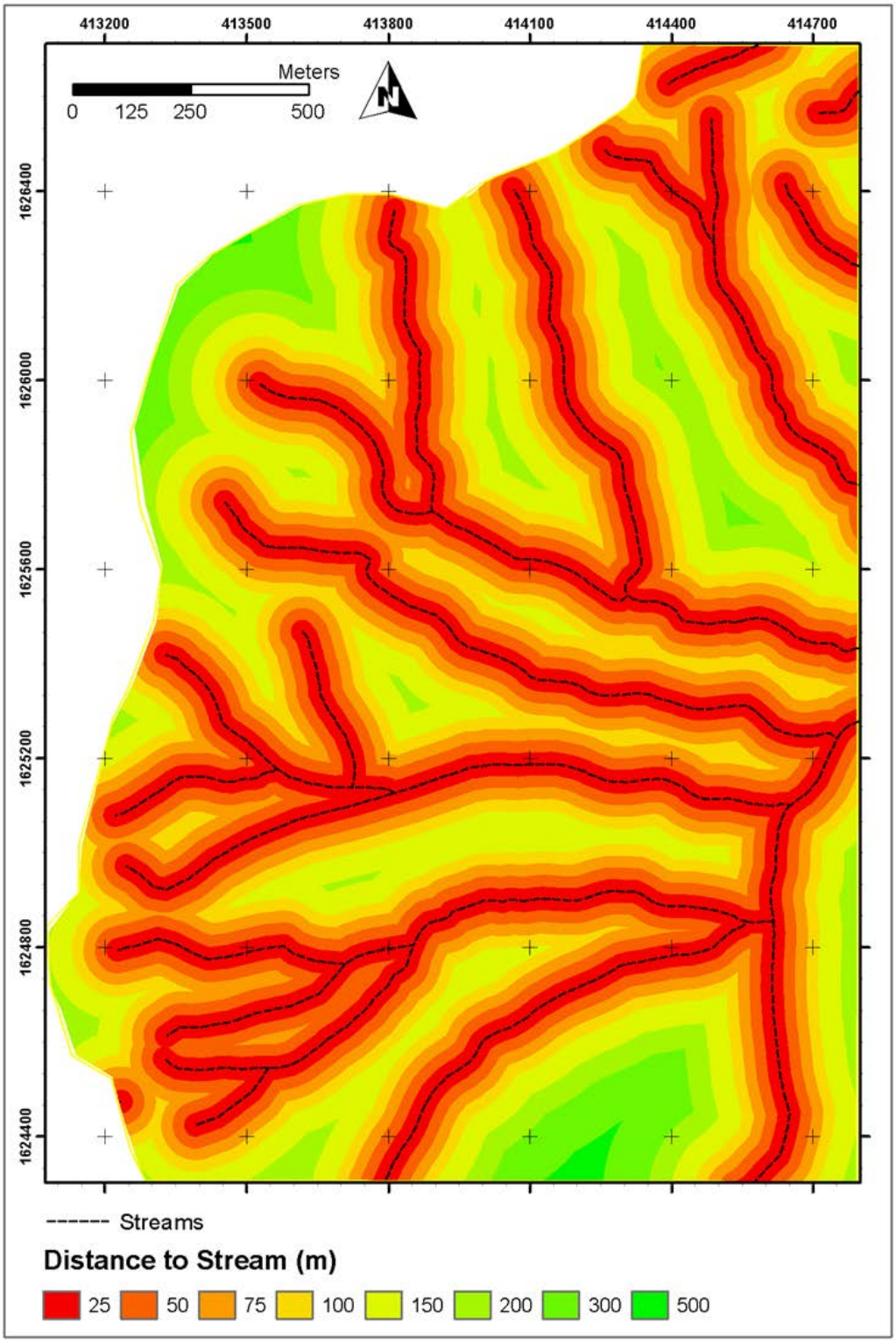


Figure 4.7: The streams of the study area buffered at increasing increments.

4.6 Distance to Fault

The study area contains 7 ligament faults and 1 circular fault from which the first 7 faults radiate (Figure 4.8). The faults were mapped by GM on their geological maps. Each fault was buffered in ArcMap in the following intervals, 0-50, 50-100, 100-200, 200-300, 300-400, 400-500, 500-600 and 600-700 meters (Figure 4.8). Like distance to streams, the ranges buffered here were chosen because these ranges were typical in similar studies and due how close the drainage channels were to each other (Chung 1999; Saha 2005; Lee 2007; Pourghasemi 2012).

4.7 Land use

The government of Guatemala developed a land use map for the whole country through the Secretary of Planning and Programming of the Presidency, referred to as Segeplan (Segeplan 2006). Five land use units cover the study area: (1) recreation, (2) corn, (3) coffee trees, (4) sparse vegetation and (5) urban areas (Figure 4.9). Sparse vegetation covers the steepest slopes in the study area (Figure 4.9).

4.8 Slope Angle

A slope map of the study area was created in ArcMap in degrees from the DEM, using ArcTool *Slope* (Figure 4.10). The steepest slopes form a ring, encasing the town (Figure 3.2 and Figure 4.10). These steep slopes have an average height of 300 meters and range from 29.9° to 62.2°. The transition away from these steep slopes to gentler slopes is stark and fast (Figure 4.10). In the dataset, the slope value of each pixel was used.

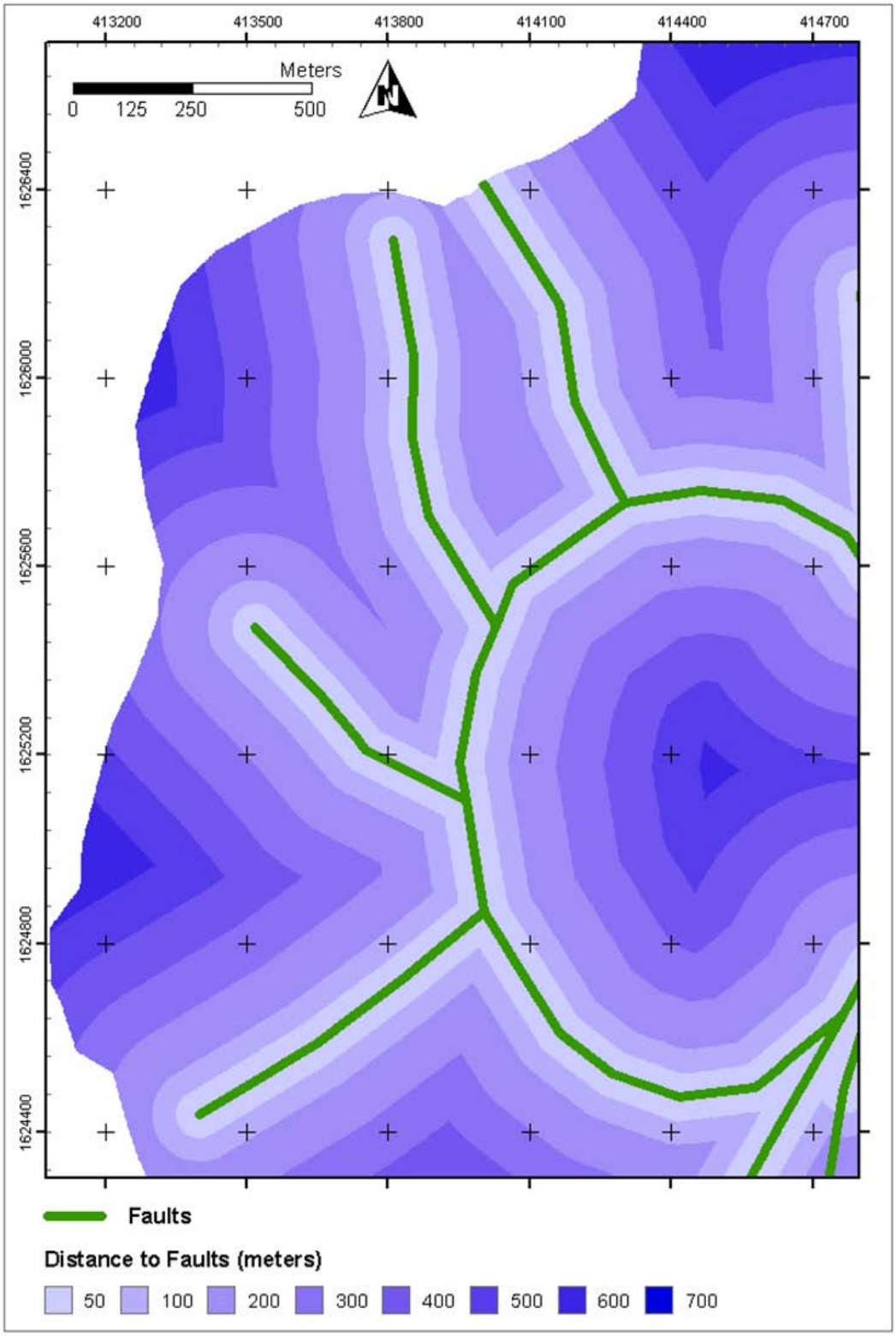


Figure 4.8: Buffered zones indicating distanct to fauts.

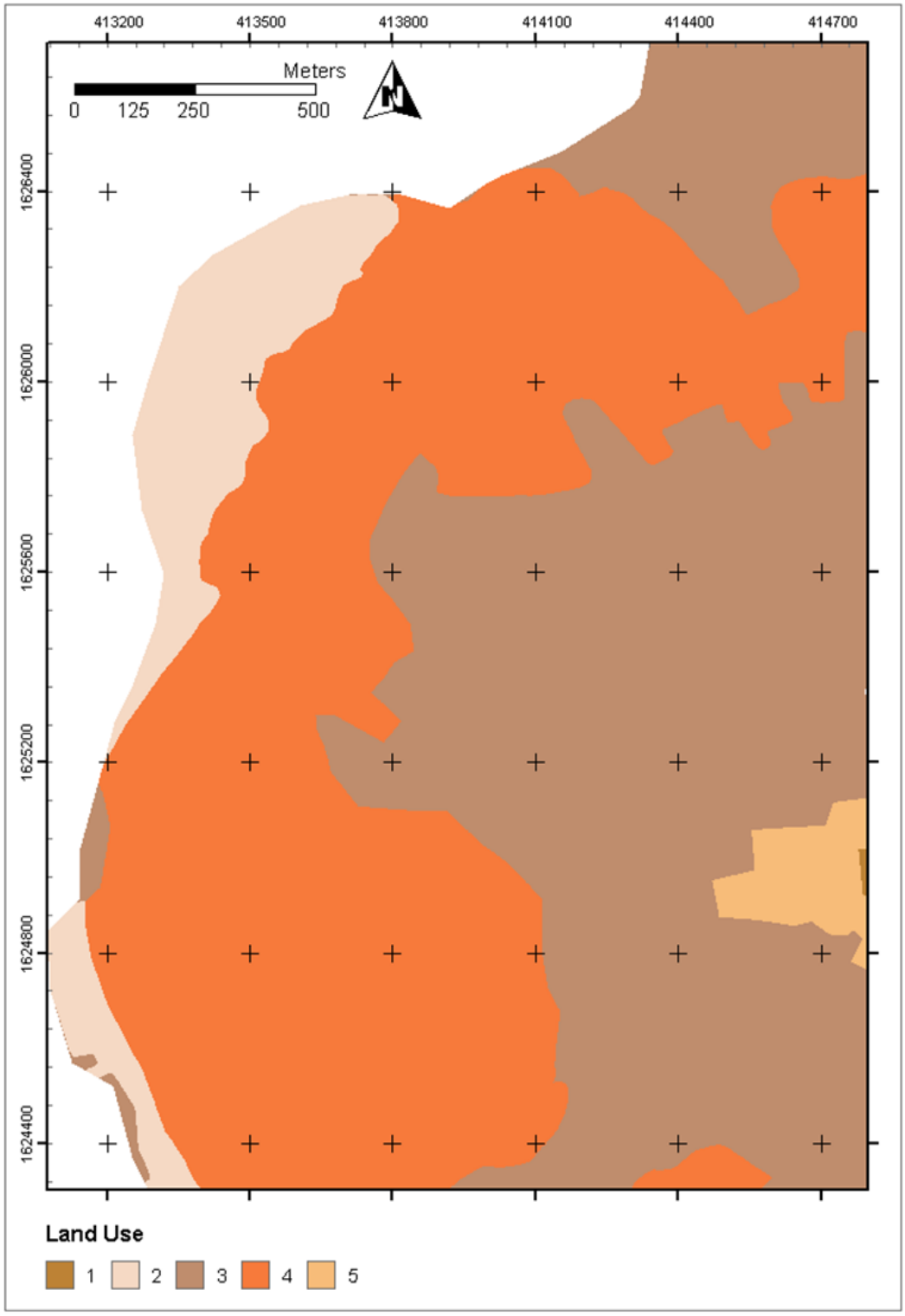


Figure 4.9: The study area overlain by the land use classifications. See Table 5.1 for explaniation of the numerical legend.

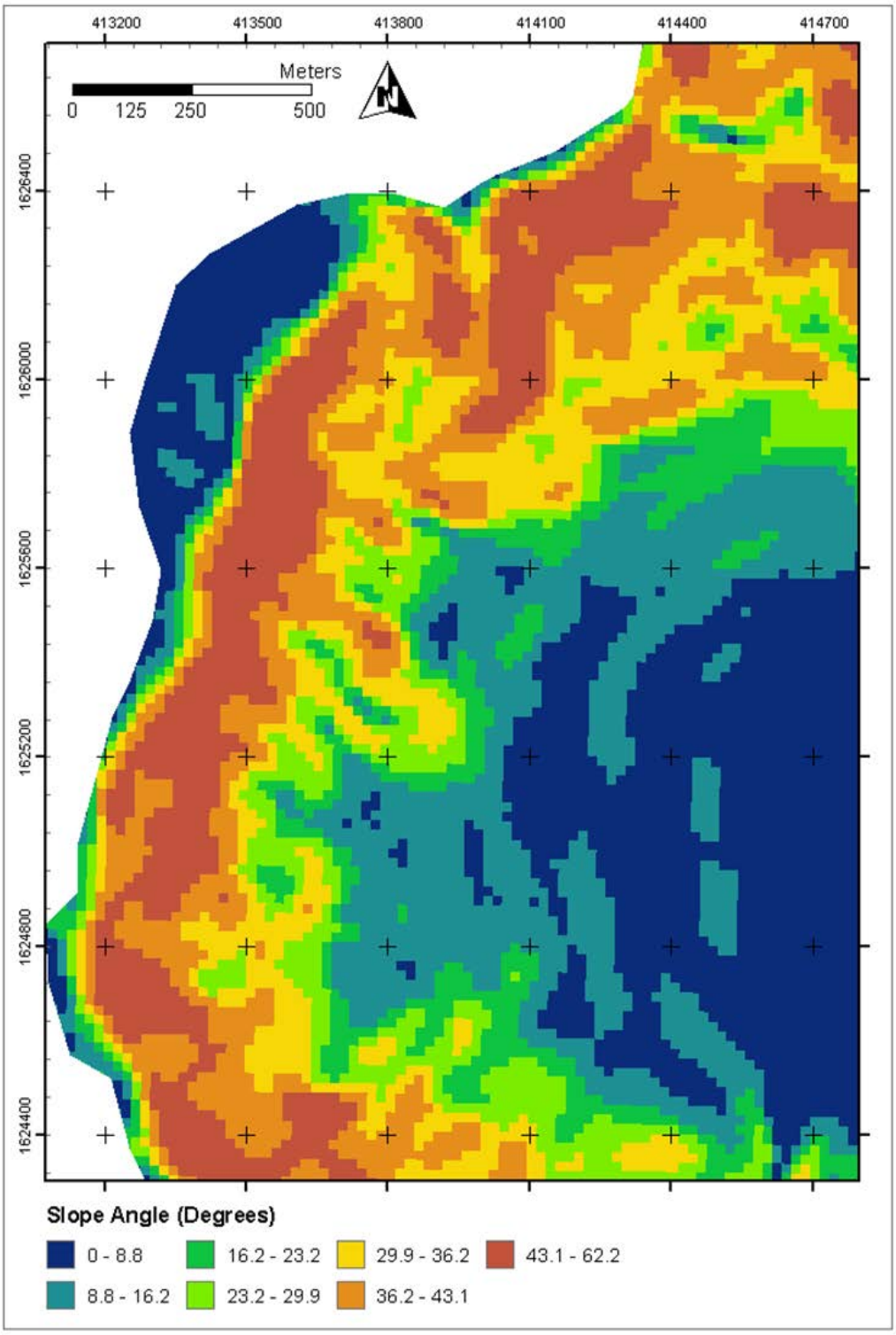


Figure 4.10: The slope raster of the study area show in degrees.

4.9 Slope Aspect

A map of slope aspects of the study area was created in ArcMap in degrees from the DEM, using ArcTool *Aspect* (Figure 4.11). The steepest slopes face in practically all the 8 cardinal directions (Figure 4.11). For the dataset individual aspect values were used.

4.10 Topographic Wetness Index (TWI)

TWI describes how topography influences the location and size of saturated areas, which act as initiation points for runoff. Wood et al. assumed steady stated conditions and uniform soil properties for their equation (1) to calculate TWI (1990).

$$(1)\ TWI = \ln(A\ /\ (\tan\ \beta))$$

A represents the specific catchments area (m^2/m) and the slope gradient is β (in degrees). Wood et al. explains that steady state and uniform soil properties can be assumed because the variation of topography often far exceeds soil transitivity.

ArcMap does not have a built in tool to evaluate the TWI. Prasad Pathak developed a python script that can be added as its own toolbox in ArcMap to evaluate TWI using the study area's DEM, DEM accuracy and cell size as inputs (Pathak 2010).

4.11 Curvature, Plan Curvature and Profile Curvature

The ArcTool *Curvature* create the rasters curvature, profile curvature and plan curvature by inputting the study area DEM. Curvature is the combination of plan and profile curvature, which separates the concavity of the slope in different directions. Profile curvature deals with how concave or convex the surface at a pixel is parallel to the maximum slope (ESRI 2012b). A negative value is upwardly convex, zero indicates a linear surface and a positive value is upwardly concave at a given cell. Plan curvature evaluates the concavity of a pixel perpendicular to the maximum slope (ESRI 2012b). A negative value equals a sidewardly concave slope, a value of zero is again linear and a positive value indicates a sidewardly convex slope. Curvature is the different possible combinations of profile and plan curvature, these combinations determine how material and water will flow over the surface (ESRI 2012b).

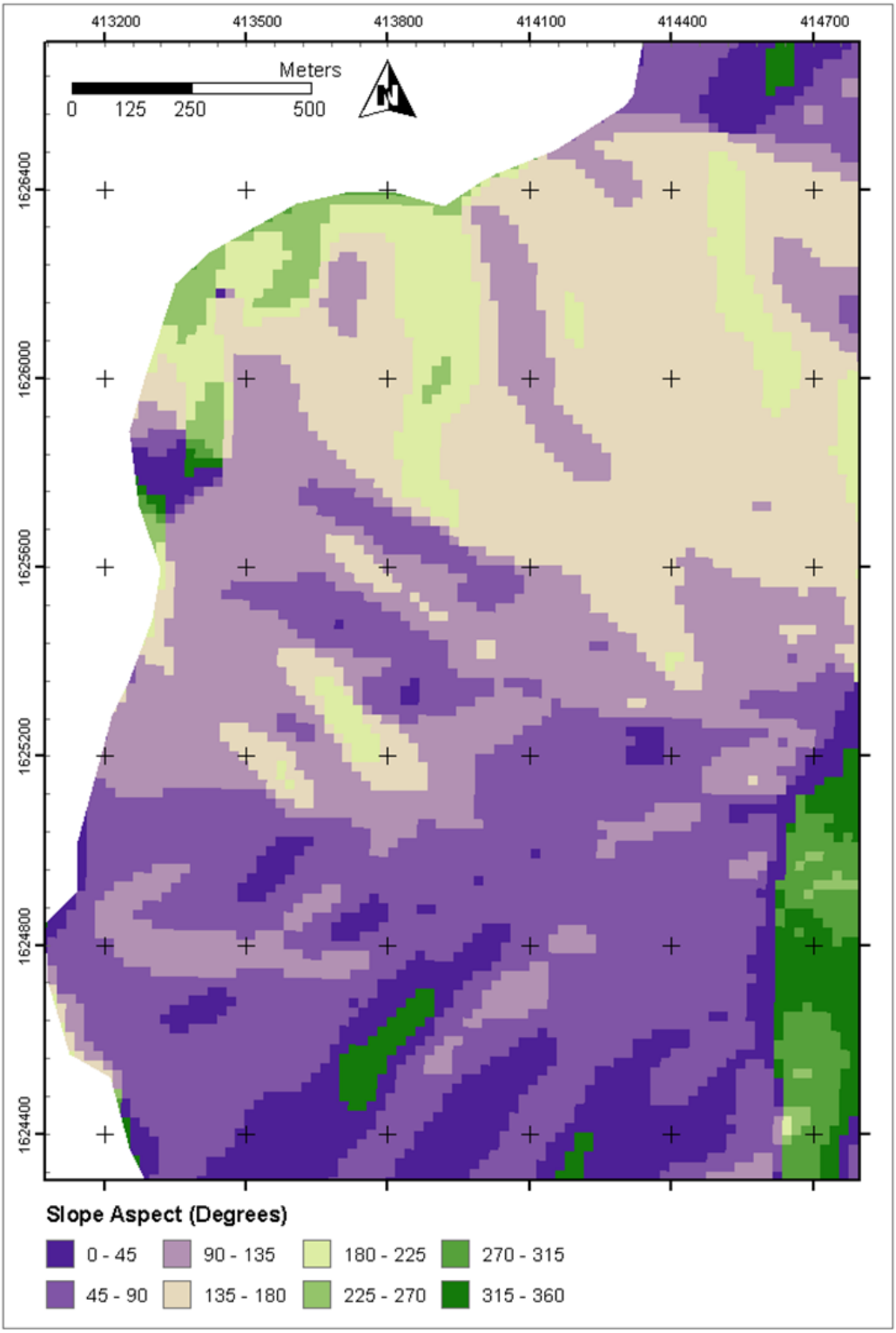


Figure 4.11: The slope aspects of the study area show in degrees.

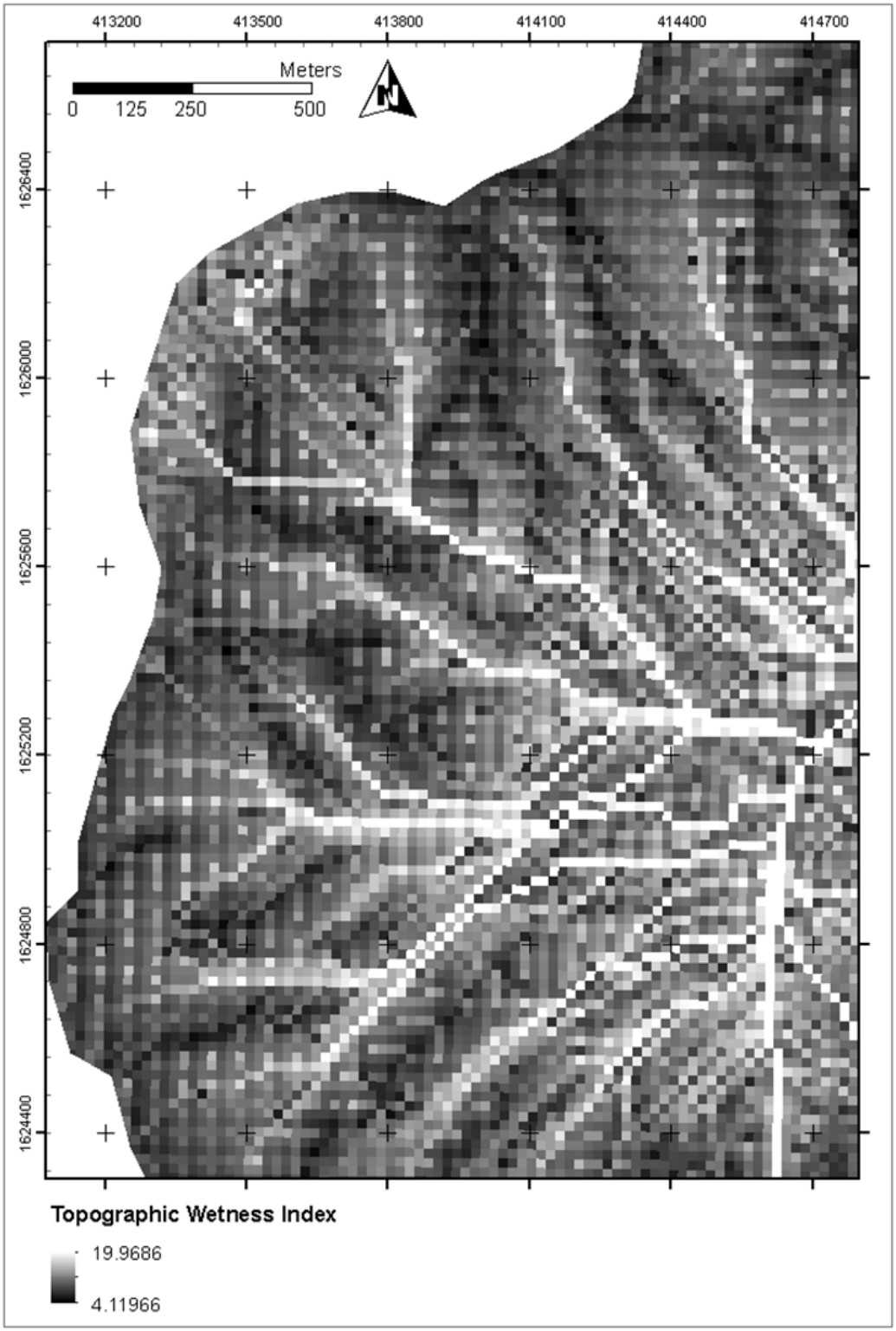


Figure 4.12: The topographic wetness index of the study area.

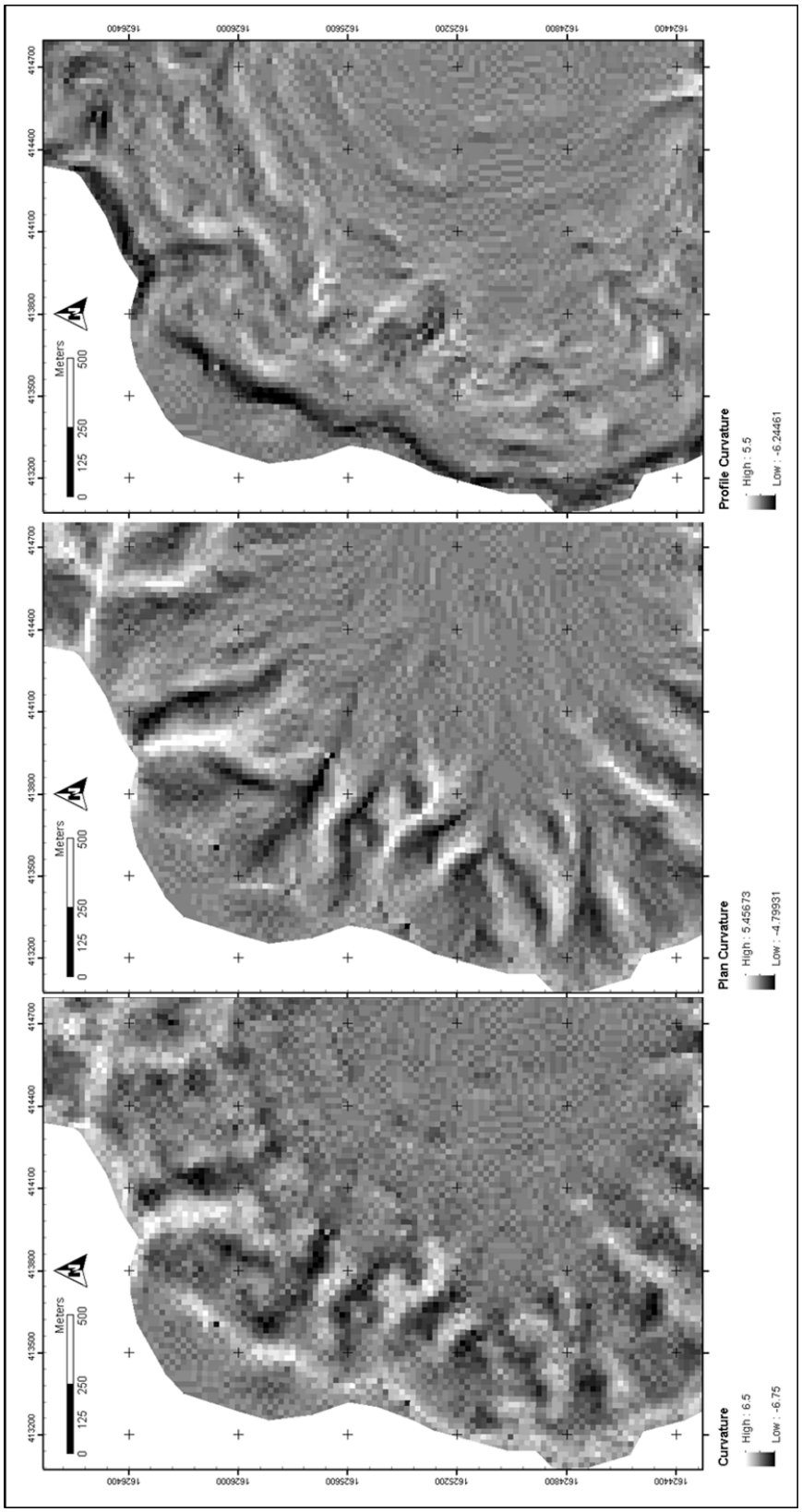


Figure 4.13: Curvature, Planform Curvature and Profile Curvature overlaying the study area.

5.0 Methodology

To use the spatial database to model for landslide probability in the region, all the primary attributes had to be rasterized. The secondary attributes were all derived from the DEM and were already rasters; the individual value of each pixel would be used in the dataset. Once all the attributes were rasterized, the data from each attribute had to be extracted to correspond with the LS/NLS database, building the data set that would train the model.

The ArcTool *Polygon to Raster* converted the primary attributes geology, geomorphology and land use to rasters. Each unit of the attributes received a unique number to distinguish that unit from the other units in the same attribute (Table 5.1).

Table 5.1: Identificaiton number for each unit in attribtues geology, geomorpholoy and land use. For descritptions of unit abbrviations see Table 4.4 and Table 4.5.

Raster Number	Geomorphology	Geology	Land Use
1	B	Qal1	Recreation
2	SAP	Tg	Corn
3	SAPD	Qcol2	Coffee Trees
4	AA	Qpf3	Sparse Vegetation
5	CLD	Qpa4	Urban Areas
6	FC	QTa	
7	PM	Qa	
8	LBCIII		

For the other primary attributes, distance to stream and distance to fault, Table 5.2 shows the numbers used to identify the buffered regions of each attribute:

Table 5.2: Identificaiton number for each buffered unit in attributes distance to streams and faults.

Distance to Streams (meters)		Distance to Faults (meters)	
Unique Raster ID	Actual Range	Unique Raster ID	Actual Range
25	0-25	50	0-50
50	25-50	100	50-100
75	50-75	200	100-200
100	75-100	300	200-300
150	100-150	400	300-400
200	150-200	500	400-500
300	200-300	600	500-600
500	300-500	700	600-700

With the LS/NLS inventory and all the primary and secondary attributes prepared as rasters, the data used to train the model, to create a probability map, was extracted. Using ArcTool *Extract Multi Value to Points*, the point of data from the same

geographical location of all the attributes were extracted for each LS and NLS point. An example of that data can be seen in Table 5.3, where N signifies a NLS point and Y a LS point.

Table 5.3: Example of complete spatial database from data set B.

LS/ NLS	Geo	GeoM	LU	Str	Flt	Slp	Asp	Curv	PlanC	ProfC	TWI
N	6	8	4	100	300	47.2	139.2	0	0.03	0.03	7.32
N	6	8	2	75	100	22.0	150.3	0	-0.54	-0.54	7.74
N	3	7	3	25	400	3.7	119.1	0.25	0.25	0	7.25
Y	4	8	3	75	500	33.4	27.1	0	-0.6	-0.6	6.35
Y	6	8	4	25	400	33.4	119.5	1.5	-0.26	-1.76	6.31
Y	6	8	4	25	400	33.4	119.5	1.5	-0.26	-1.76	6.31

A complete table of all the attribute information corresponding to each LS and NLS point in data set A and B can be found in appendix 2.

To develop the ideal model certain combinations of available attributes will provide better results over others. Weka (Hall et al. 2009) provides various attribute analyzers to help gauge the importance of each attribute, including information gain (IG), gain ratio (GR), chi-squared (CS) and filtered subset evaluation (FS) (Hall et al. 2009). Miner et al. (2010) used information gain (IG) on the Bellarine Peninsula in Victoria, Australia for a landslide susceptibility assessment; Marjanovic et al. (2011) and Tein Bui et al. (2012) used gain ratio in a landslide susceptibility assessment in the Starča Basin, Croatia and in Vietnam respectively. Hwang et al. (2009) used IG, GR, CS and others for slope failure prediction in South Korea. Hwang et al. (2009) noted different rankings of the attributes in the attribute analyzers they used, ultimately averaging all the results for final rankings of importance. No instance of FS being using for data mining LS causative attributes was found, but the method has been applied in other fields (Mujalli et al. 2011; Bharti et al. 2010).

The following analyses were carried out on the attributes for the study area: IG, GR, CS and FS. IG evaluates an attribute by determining the overall information gain with respect to the class (Witten et al. 2005), meaning attributes with the most information are given higher ranks (Hwang et al. 2009). In other words, the attributes that have seemingly more control on whether an instance is LS or NLS is given a higher

rank. GR evaluates each attribute similarly to IG, but has less bias towards attributes that have a wide range (Hwang et al 2009). CS calculates the relevance of an attribute through its chi-squared statistic with respect to the occurrence of a landslide, or not (Witten et al. 2005; Hwang et al 2009). The chi-squared statistic starts by separating each instance, in this case LS and NLS, with respect to the attribute in question, and then deciding if adjacent instances should be merged (Witten et al. 2005). Deciding if instances should be merged relies on a confidence interval being met, X^2 – the chi-squared statistic (Witten et al. 2005). This test of meeting the confidence interval is repeated until no other merges are possible (Witten et al. 2005). FS evaluates subsets of the attributes by examining intrinsic characteristics of the attribute data (Mujalli et al. 2011). IG, GR and CS give each attribute a rank of importance based on the amount of information gained by a given attribute. FS reports how many times an attribute was folded, the more folds indicate a more influential attribute.

All four attribute analyzers were conducted in a supervised fashion, each instance was known whether the data corresponded to a LS or NLS point and a 10-fold cross validation was used. A cross fold validation takes the number of available instances in a data set and divides the data into 10 random data subsets using a seed number to determine the cutoff (1 was used in this study). Miner et al. (2010) observed no difference between supervised and unsupervised usage of IG in their landslide hazard assessment.

Once the attributes had been evaluated, the whole dataset was brought into the *classify* tab in Weka (Hall et al. 2009). Multiple algorithms were tested from those available in Weka (Hall et al. 2009) to see which yielded the best initial results. Each algorithm test was supervised by a 10-fold cross validation. A series of statistical measures determined the initial success of each algorithm tested.

Weka (Hall et al. 2009) automatically calculates various statistical measures to evaluate the accuracy of each model, including kappa coefficient, root mean squared, true positive rate (TPR), false positive rate (FPR), the confusion matrix, precision, recall, relative operating characteristics (ROC) curve area, and F measure (Figure 5.1). Overall accuracy can also be calculated from the results in the confusion matrix.

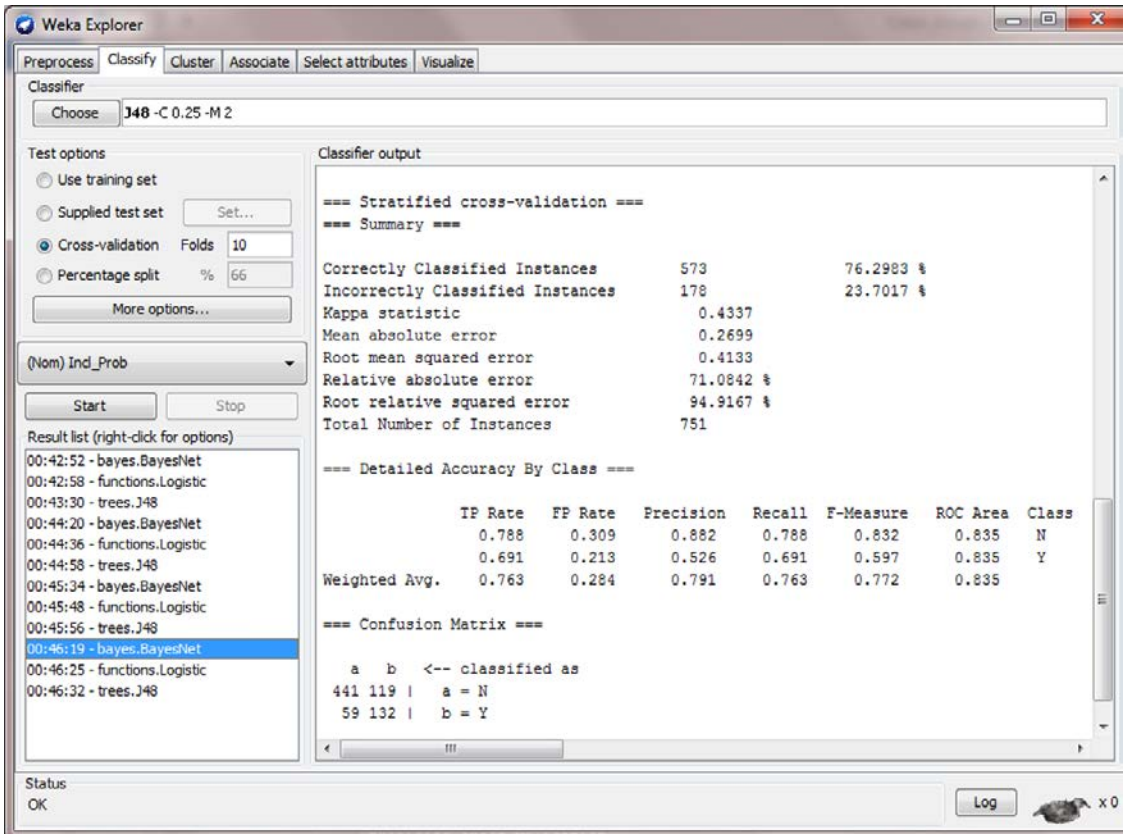


Figure 5.1: Example of statistical measures calculated by Weka.

The statistical measures optimized to choose the best algorithms to test in this study were TPR (recall), precision, the confusion matrix, overall accuracy, F measure, and ROC area.

Oommen et al. (2010) describes TPR, also known as recall, to be instances when known points of LS are predicted to be LS by the model, or when known NLS points are predicted to be NLS (Equation 5.1). Recall evaluates classes of observations, in the context of this study: LS or NLS points. For example, a recall of 1 for LS points means every instance of known LS was predicted to be LS, however, this does not include misclassified instances of NLS.

$$recall = TPR = TP/(TP + FN) \qquad \text{(Equation 5.1)}$$

Where TP is the total number of LS instances correctly identified as LS; and FN is the total number of actual LS instances that the model classified as NLS (Oommen et al. 2010).

Oommen et al. (2010) describes precision to be the predicted instances, either LS or NLS, which match what the instance was known to have experienced (Equation 5.2). For example, a precision of 1 for LS points indicates that every instance predicted to have a LS had a LS, however, this does not include misclassified instances of LS.

$$precision = TP/(TP + FP) \qquad \text{(Equation 5.2)}$$

Where FP is the total number of NLS instances that the model classified as LS (Oommen et al. 2010).

The confusion matrix compares the known instances of LS and NLS with the model predictions of each instance (Oommen et al. 2010) (Table 5.4).

Table 5.4: The confusion matrix displaying known instances of LS and NLS versus how the model classifies, predicted instances, each instance based on the data of the attributes. *Table modified from Ooommen et al. 2010.*

Confusion Matrix		Known Instances	
		NLS	LS
Predicted Instances	NLS	TP	FP
	LS	FN	TN

Overall Accuracy (Equation 5.3) of the model can be calculated from the confusion matrix using TP, TN, FP and FN:

$$overall\ accuracy = (TP + TN)/(TP + TN + FP + FN) \qquad \text{(Equation 5.3)}$$

F measure (Equation 5.4) combines precision and recall (Hall et al. 2009):

$$\text{F measure} = 2 * \text{Precision} * \text{Recall}/(\text{Precision} + \text{Recall}) \qquad \text{(Equation 5.4)}$$

F measure is the harmonic mean of precision and recall.

Weka (Hall et al. 2009) uses ROC area, or area under the curve (AUC), to help indicate a model's success, where an ROC of 1 indicates all the known instances of both classes, LS and NLS, were correctly predicted (Yesilnacar et al. 2005).

Recall, precision, the confusion matrix, overall accuracy, F measure, and ROC area helped isolate the algorithms that were superior to other available options in Weka (Hall et al. 2009). Decision Tree J48 (J48), logistic regression and BayesNet gave the best initial results.

In Weka (Hall et al. 2009) J48 is a decision tree that uses the algorithm C4.5, originally developed by J. Ross Quinlan in the early 1970s (Witten et al. 2005). Decision

trees work from a top-down approach (Witten et al. 2005), splitting data sets into consecutively smaller groups depending on the data of an instance for one of its attributes (Hwang et al. 2009) as seen in Figure 5.2. Figure 5.2 is an example of a J48 output from this data set. The decision tree in Figure 5.2 starts with geomorphology, and the first split is based on whether instances, of LS or NLS, are found with raster identification numbers of less than or equal to 7 or greater than 7. For the instances found in the geomorphological sections 7 or under (see Table 5.1 and Figure 4.6) the tree predicted they were NLS, of the 280 instances predicted to NLS, 3 were misclassified (Figure 5.2). For the instances found in the geomorphological sections over 7, therefore section 8, the next split was based on distance to faults (Figure 5.2). The decision tree keeps creating distinctions based on the available attributes until all the instances of the data set are resolved.

The second algorithm that showed promise for LS prediction was logistic regression, in this case, binary logistic regression, because only two discrete dependent variables LS or NLS are possible. In its simplest form the regression is written:

$$P = 1 \div (1 + e^{-z}) \quad \text{(Equation 5.5)}$$

Where *P* equals the probability, from 0 to 1, of a landslide occurring. And z equals:

$$z = B_0 + B_1X_1 + B_2X_2 + \ldots + B_nX_n \quad \text{(Equation 5.6)}$$

Where B_0 equals the y-intercept of the model; B_i (i = 1, 2 … n) equals the slope coefficient of the model; and X_i (i = 1, 2 … n) is the independent variable being included in the model iteration. The value of *n* in any model iteration depends on the combination of independent variables, attributes, being included.

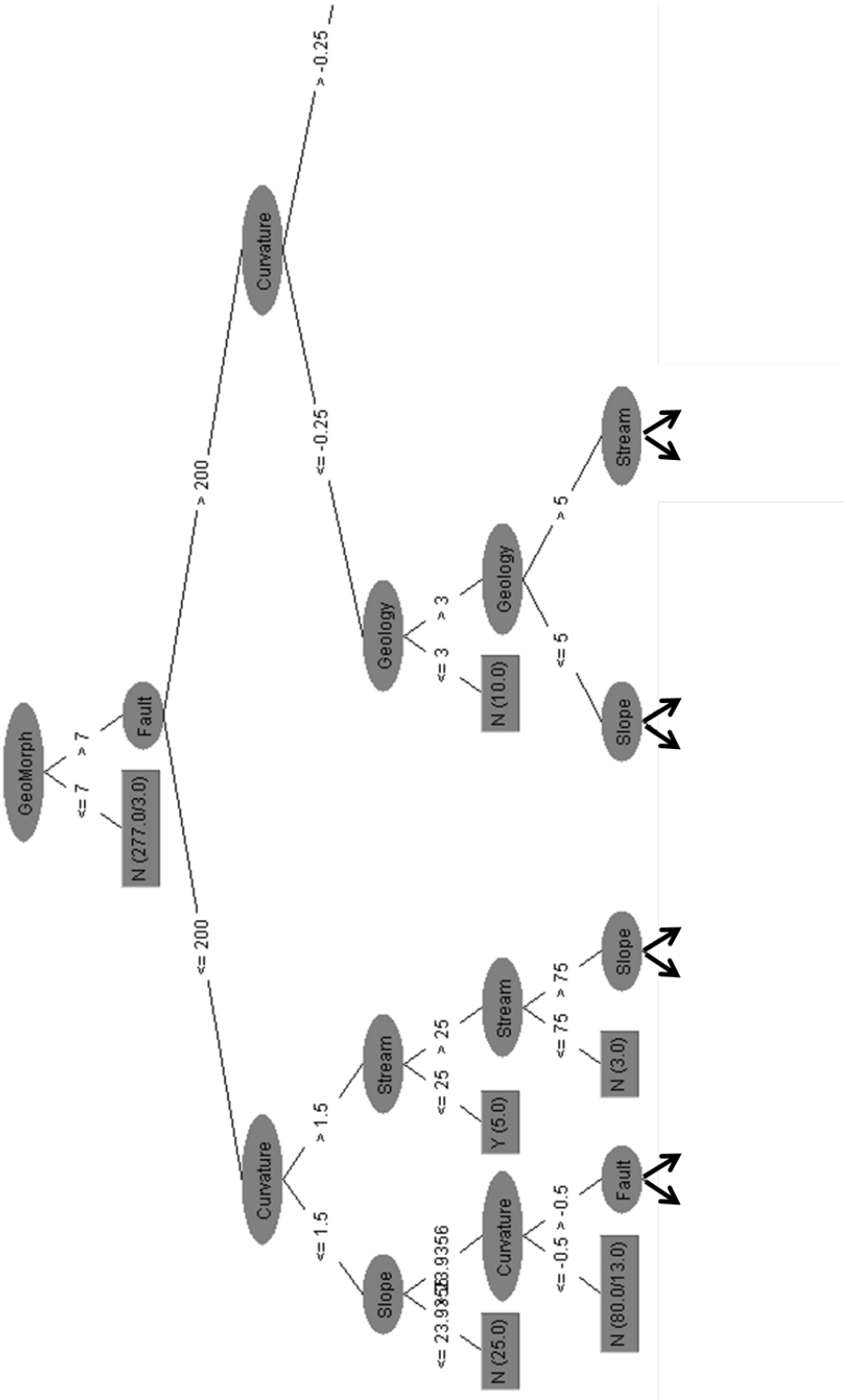


Figure 5.2: The top portion of a decision tree using Data Set B and the following attribtues: distance to faults and streams, geology, geomorphology, slope and curvature.

The last algorithm that yielded good preliminary results was BayesNet. BayesNet learns Bayesian networks under two assumptions: 1.) all attributes are pre-discretized and 2.) there are no missing values. While both decision trees and Bayesian networks can determine probability of an event occurring, decision trees divide the data set into successively smaller pieces, which can yield less reliable probability estimates (Witten et al. 2005). Bayesian networks address that by employing directed acyclic graphs, which is a network of nodes. Each attribute gets a node; they are then interconnected so there are no cycles (Witten et al. 2005). Within each node are defined probability distributions (Witten et al. 2005) used to predict the probability of a LS or NLS.

Once J48, logistic regression and BayesNet were identified, each algorithm was tested multiple times. Each iteration of the three algorithms used with different attribute combinations, using the attribute analyzer results as a guide in choosing the most landslide causative attributes. All algorithms were supervised with a 10-fold cross validation. Once the ideal model was identified, using data set B, the same model was used on data set A to compare the success of each model.

To create the susceptibility map, 10,000 random points were chosen in the study area using the ArcTool *Create Spatially Balance Points*. To control which points were chosen, the *Environments Settings* tab on the ArcTool was used to choose the seed number. The seed number is controlled under the option *Random Numbers*. The seed number used was 89 with *Random Generator Type* ACM599. Then at each of the 10,000 points the corresponding pixel values from the attributes were extracted using ArcTool *Extract Multi Value to Points*, creating data set C. Dataset C was then imported into Weka and evaluated using the model results from the ideal algorithm, yielding a percent likelihood of being a LS point.

The probability of being a LS in dataset C was imported into ArcMap to perform inverse distance weighted (IDW) spatial interpolation to create a probability map for the study area. IDW interpolates an unknown cell's value by combining, linearly, known values of other local points, weighting each of the known values based on how close they are to the unknown cell; points that are closer are given more weight, importance in determining the value of the unknown cell.

After interpolating and creating the probability map, the divisions of the probability percentages were determined in part by many studies equal division from 0 to 100 (Suzen et al. 2004; Yesilnacar et al. 2005; Ozdemir 2011) and that the GM study ultimately made their susceptibility with 5 risk rankings: very low, low, moderate, high and very high. Suzen et al. 2004, Yesilnacar et al. 2005 and Ozdemir 2011 dealt with the arbitrary binning of probability by equal intervals by calculating the area each zone covered in the study area and the number of landslides present in each zone from the data set that was used to develop the probability map. Further, exactly how the divisions were broken, on a scale of 0 to 100, depended on the highest susceptibility probability in the study area. To be able to draw more comparisons between this study and GM's study (2012a) five divisions were chosen: 0-20%, 20-40%, 40-60%, 60-80% and 80-100%

To independently validate the susceptibility map, landslides from Tropical Storm Agatha were superimposed on the map. A successful validation would have the majority of initiation zones of the Agatha landslides in the areas with the highest susceptibility.

6.0 Results and Discussion

IG, CS and GR ranked importance from 1 to 11, based on the total number of attributes, with one being the most important. FS reported how many times each attribute was folded. The attribute analyzers, IG, CS and GR, ranked the most important attributes to the least important, from 1 to 11 (Table 6.1). Since a 10-fold cross validation was used for FS, the attributes that had 10 folds meant they were the most influential in determining LS/NLS separation (Table 6.1).

IG and CS had the same ordering of attributes in their rankings, though their average ranks for each attribute differed slightly depending on how the attribute was used in the 10-fold cross validations. GR was very similar to IG and CS, but it ranked geomorphology, slope and plan curvature higher relative to the other attributes (Table 6.1). FS found TWI, distance to streams, land use, plan curvature, curvature and aspect to have no folds, where IG, CS and GR ranked them at five or higher (Table 6.1). One key difference is IG, CS and GR found distance to faults to have an average rank of 7, where FS found distance to faults to be the fourth most LS indicative with 2 folds (Table 6.1).

GR and FS found the top three attributes to be geomorphology, slope and profile curvature (Table 6.1). Another key difference between the four evaluators is FS had geology as the fifth most important attribute, where IG and CS found it to be number one and GR found it to be the fourth (Table 6.1).

Table 6.1: Attribute analysis results using difference evaluators.

IG		**CS**		**GR**		**FS**	
Average Rank		*Average Rank*		*Average Rank*		*# of Folds*	
1.1	Geology	1.2	Geology	1	Geomorph	10	Geomorphology
1.9	Geomorph	1.9	Geomorph	2	Slope	10	Slope
3	Slope	3.1	Slope	3	Prof. Curvature	10	Prof. Curvature
4	Prof. Curvature	3.8	Prof. Curvature	4	Geology	2	Fault
5.4	Curvature	5.1	Curvature	5	Curvature	1	Geology
5.8	TWI	6	TWI	6.1	TWI	0	TWI
6.9	Fault	6.9	Fault	7	Fault	0	Stream
7.8	Land use	8.1	Land use	8.2	Land use	0	Land Use
8	Stream	9.2	Stream	9.3	Plan Curvature	0	Plan Curvature
10	Plan Curvature	10	Plan Curvature	10	Aspect	0	Curvature
10.7	Aspect	10.7	Aspect	10.3	Stream	0	Aspect

The attribute analysis results in Table 6.1 were used as a guide to process data set B with the three algorithms: decision tree J48, logistic regression and BayesNet. Each model was processed multiple times using different attribute combinations to obtain the best model for each algorithm. Choosing which attributes to include and eliminate was informed by the attribute analysis results (Table 6.1) and by evaluating the statistical parameters.

Table 6.2, Table 6.3 and Table 6.4 show the first model iteration of each algorithm (J48-1, Log-1 and Bay-1), which included all the attributes. Following the first iteration, the tables show attributes being eliminated. In the second iteration of each model one to three attributes are shown to have been eliminated (Table 6.2, Table 6.3 and Table 6.4). If the elimination of one attribute, as shown in Table 6.2, yielded significant improvement in a new model that model was included in the table. For logistic regression and BayesNet, the elimination of one attribute did not contribute to model improvement as compared to the model including all the attributes, therefore the second model included in their respective tables had more attributes eliminated (Table 6.3 and Table 6.4).

The order the attributes were eliminated, for all three algorithms, were preferentially based on the results of the attribute analyzers (Table 6.1). The second

decision tree model, J48-2, had only aspect eliminated, the attribute that ranked the lowest in IG and CS, and no folds in FS (Table 6.1 and Table 6.2). The logistic regression saw improvement in model Log-2 with the removal of aspect and plan curvature, two attributes the attribute analysis results found to have little influence (Table 6.1 and Table 6.3). The BayesNet saw improvement in model Bay-2 with the removal of aspect, plan curvature and distance to stream, again, attributes that all ranked low in the attribute analysis results (Table 6.1 and Table 6.3). The attribute analyzer results (Table 6.1) guided the attributes that were chosen for removal, especially in the first iterations of each algorithm.

After the removal of the lowest ranking attributes, plan curvature and slope aspect, the ones that were removed for each algorithm depended more on trial and error. For instance the ideal model for J48 included distance to streams in J48-4 (Table 6.2). Model J48-5 eliminated distance to streams one of lowest ranking attributes according to Table 6.1, but it contributed to the degradation of that model iteration. While the attribute analyzer results guided which attributes were preferentially eliminated, the tables below do not necessarily show attribute eliminations that strictly follow Table 6.1. If an attribute was eliminated and the model iteration performed resulting in significant degradation that attribute was re-added. Following the readdition of the previously eliminated attribute another attribute(s) was eliminated for the next model iteration, and so forth. Not all iterations tried with the three algorithms are shown, some models were redundant and unnecessary to show. The model iterations for each algorithm shown in Table 6.2, Table 6.3 and Table 6.4 are meant to show the baseline model with all the attributes, model optimization with a select number of attributes excluded and model degradation indicating the attributes that were integral to the success of the model.

Table 6.2: Decision Tree J48 Models. Each model is a different combination of attributes, the grayed and striked text indicats attributes that have been removed. The highest number of each parameter, across all models, has red text and the second highest parameter is shown in bold. The best model is outlined in blue. KNLS and KLS stand for known NLS and LS points; PNLS and PLS stand for predicted NLS and LS points.

J48 Models:		**J48-1**		**J48-2**		**J48-3**		**J48-4**		**J48-5**		**J48-6**		**J48-7**		**J48-8**		**J48-8**	
		KNLS	KLS	KNLS	KLS	KNLS	KLS	KNLS	KLS	KNLS	KLS	KNLS	KLS	KNLS	KLS	KNLS	KLS	KNLS	KLS
Confusion Matrix	PNLS	498	62	502	58	499	61	505	55	503	57	498	62	490	70	471	89	553	7
	PLS	96	85	96	95	90	101	89	102	100	91	106	85	106	85	81	110	180	11
F-Measure of LS		0.546		0.552		**0.572**		**0.586**		0.537		0.503		0.491		0.564		0.105	
F-Measure of NLS		0.863		0.867		0.869		**0.875**		0.865		0.856		0.848		**0.874**		0.855	
Recall (TPR) of LS		0.497		0.497		0.529		**0.534**		0.476		0.445		0.445		**0.576**		0.058	
Recall (TPR) of NLS		0.889		0.896		0.891		**0.902**		0.898		0.889		0.875		0.841		**0.988**	
Precision of LS		0.605		0.621		**0.623**		**0.650**		0.615		0.578		0.548		0.553		0.611	
Precision of NLS		0.838		0.839		0.847		**0.850**		0.834		0.825		0.822		**0.853**		0.754	
Correctly Identified*		78.96%		79.50%		**79.89%**		**80.83%**		79.10%		77.63%		76.57%		77.36%		75.10%	
Overall Accuracy		0.693		0.697		**0.710**		**0.718**		0.687		0.667		0.660		0.709		0.523	
ROC Area		0.755		0.768		0.783		0.792		0.769		0.788		**0.804**		**0.802**		0.745	
Attributes	1	Geology		Geology		Geology		Geology		Geology		Geology		Geology		Geology		Geology	
	2	GeoMorph		GeoMorph		GeoMorph		GeoMorph		GeoMorph		GeoMorph		GeoMorph		GeoMorph		GeoMorph	
	3	Streams		Streams		Streams		Streams		~~Streams~~		~~Streams~~		~~Streams~~		~~Streams~~		~~Streams~~	
	4	Faults		Faults		Faults		Faults		Faults		Faults		~~Faults~~		~~Faults~~		~~Faults~~	
	5	Land use		Land use		Land use		Land use		Land use		~~Land use~~		~~Land use~~		~~Land use~~		~~Land use~~	
	6	Slope		Slope		Slope		Slope		Slope		Slope		Slope		Slope		Slope	
	7	Aspect		~~Aspect~~		~~Aspect~~		~~Aspect~~		~~Aspect~~		~~Aspect~~		~~Aspect~~		~~Aspect~~		~~Aspect~~	
	8	TWI		TWI		TWI		TWI		TWI		TWI		TWI		TWI		TWI	
	9	Curvature		Curvature		Curvature		Curvature		Curvature		Curvature		Curvature		~~Curvature~~		~~Curvature~~	
	10	Plan Curv.		Plan Curv.		~~Plan Curv.~~		~~Plan Curv.~~		~~Plan Curv.~~		~~Plan Curv.~~		~~Plan Curv.~~		~~Plan Curv.~~		~~Plan Curv.~~	
	11	Prof. Curv.		Prof. Curv.		Prof. Curv.		~~Prof. Curv.~~		Prof. Curv.		Prof. Curv.		Prof. Curv.		Prof. Curv.		~~Prof. Curv.~~	

Between models J48-1 through J48-7 all the key statistical parameters (recall, precision, confusion matrix, F-measure, ROC and overall accuracy) improve until model J48-4, after which they begin to degrade, with the exception of the ROC of J48-7, until the last model (Table 6.2). While J48-6 has the highest ROC area, all the other measures are inferior to the other models (Table 6.2). Model J48-4 has the superior F-measure of LS and NLS, precision of LS, percentage correctly identified instances and overall accuracy, and the second best results for recall of LS and NLS and precision of NLS. The only parameter that J48-4 has neither the superior or second highest result is ROC area, J48-7 has the highest ROC area of 0.804. While J48-8 has the highest recall of LS points, the recall of NLS is the lowest out of all the model iterations (Table 6.2). J48-4, which excludes attributes slope aspect, plan curvature and profile curvature, is the superior decision tree model, having the most optimized statistical measures (Table 6.2).

Unlike with the J48 models that display a stark peak, improvement with select attribute elimination and degradation when too many were eliminated, the results of the logistic regression models vacillate as different combinations of attributes are eliminated (Table 6.3).

Table 6.3: Multiple Logistic Regression Models. Each model is a different combination of attributes, the grayed and striked text indicats attributes that have been removed. The highest number of each parameter, across all models, has red text and the second highest parameter is shown in bold. The best model is outlined in blue. KNLS and KLS stand for known NLS and LS points; PNLS and PLS stand for predicted NLS and LS points. In Log-2 both aspect and plan curvature have been removed, because removing both yielded the same results as removing just aspect alone. The combination of removing all the attributes seen in Log-5 plus TWI is not shown becauase the results were nearly the same as model Log-5.

Logistic Models:		**Log-1**		**Log-2**		**Log-3**		**Log-4**		**Log-5**		**Log-6**		**Log-7**		**Log-8**	
		KNLS	KLS	KNLS	KLS	KNLS	KLS	KNLS	KLS	KNLS	KLS	KNLS	KLS	KNLS	KLS	KNLS	KLS
Confusion Matrix	PNLS	513	47	511	49	511	49	514	46	524	36	511	49	513	47	505	55
	PLS	105	86	105	86	107	84	104	87	114	77	102	89	101	90	118	73
F-Measure of LS		0.531		0.528		0.519		0.537		0.507		**0.541**		**0.549**		0.458	
F-Measure of NLS		0.871		0.869		0.868		0.873		**0.875**		0.871		**0.874**		0.854	
Recall (TPR) of LS		0.450		0.450		0.440		0.455		0.403		**0.466**		**0.471**		0.382	
Recall (TPR) of NLS		0.916		0.913		0.913		**0.918**		**0.936**		0.913		0.916		0.902	
Precision of LS		0.647		0.637		0.632		0.654		0.681		0.645		**0.657**		0.570	
Precision of NLS		0.830		0.830		0.827		0.832		0.821		**0.834**		**0.836**		0.811	
Correctly Identified*		79.76%		79.49%		79.23%		80.03%		**80.03%**		79.89%		**80.29%**		76.96%	
Overall Accuracy		0.683		0.682		0.677		**0.687**		0.670		0.685		**0.694**		0.642	
ROC Area		0.839		0.840		0.841		0.842		0.827		**0.844**		**0.846**		0.809	
Attributes	1	Geology		Geology		Geology		Geology		Geology		Geology		Geology		Geology	
	2	GeoMorph		GeoMorph		GeoMorph		GeoMorph		GeoMorph		GeoMorph		GeoMorph		GeoMorph	
	3	Streams		Streams		~~Streams~~		~~Streams~~		~~Streams~~		~~Streams~~		~~Streams~~		~~Streams~~	
	4	Faults		Faults		Faults		Faults		~~Faults~~		Faults		Faults		Faults	
	5	Land use		Land use		Land use		~~Land use~~		~~Land use~~		~~Land use~~		~~Land use~~		~~Land use~~	
	6	Slope		Slope		Slope		Slope		Slope		Slope		Slope		Slope	
	7	Aspect		~~Aspect~~		~~Aspect~~		~~Aspect~~		~~Aspect~~		~~Aspect~~		~~Aspect~~		~~Aspect~~	
	8	TWI		TWI		TWI		TWI		TWI		~~TWI~~		~~TWI~~		~~TWI~~	
	9	Curvature		Curvature		Curvature		Curvature		Curvature		Curvature		~~Curvature~~		~~Curvature~~	
	10	Plan Curv.		~~Plan Curv.~~		~~Plan Curv.~~		~~Plan Curv.~~		~~Plan Curv.~~		~~Plan Curv.~~		~~Plan Curv.~~		~~Plan Curv.~~	
	11	Prof. Curv.		Prof. Curv.		Prof. Curv.		Prof. Curv.		Prof. Curv.		Prof. Curv.		Prof. Curv.		~~Prof. Curv.~~	

Of the logistic regression models, Log-7 had the best combination of recall, precision and F-measure for the LS and NLS points, and the best results for correctly identified instances, overall accuracy and ROC (Table 6.3). Log-7 did not have the highest recall for NLS points, which is found in Log-5, but overall accuracy and ROC were the highest of all the models in model Log-7 (Table 6.3). The model began to degrade with the removal of Profile Curvature in Log-8 (Table 6.3).

Table 6.4: Multiple Bayesian Network Models. Each model is a different combination of attributes. The highest number of each parameter, across all models, has red text and the second highest parameter is shown in bold. The best model is outlined in blue. KNLS and KLS stand for known NLS and LS points; PNLS and PLS stand for predicted NLS and LS points.

Bayesian Models		**Bay-1**		**Bay-2**		**Bay-3**		**Bay-4**		**Bay-5**		**Bay-6**		**Bay-7**	
		KNLS	KLS	KNLS	KLS	KNLS	KLS	KNLS	KLS	KNLS	KLS	KNLS	KLS	KNLS	KLS
Confusion Matrix	PNLS	413	147	416	144	441	119	379	181	436	124	443	117	337	223
	PLS	46	145	48	143	59	132	40	151	49	142	53	138	29	162
F-Measure of LS		0.600		0.598		0.597		0.577		**0.621**		**0.619**		0.563	
F-Measure of NLS		0.811		0.813		0.832		0.774		**0.834**		**0.839**		0.728	
Recall of LS		0.759		0.749		0.691		**0.791**		0.743		0.723		**0.848**	
Recall of NLS		0.738		0.743		0.788		0.677		**0.779**		**0.791**		0.602	
Precision of LS		0.497		0.498		0.526		0.455		**0.534**		**0.541**		0.421	
Precision of NLS		**0.900**		0.897		0.882		**0.905**		0.899		0.893		0.921	
Correctly Identified*		74.30%		74.43%		76.30%		70.57%		**76.96%**		**77.36%**		66.44%	
Overall Accuracy		0.749		0.750		0.740		0.734		**0.761**		**0.757**		0.725	
ROC Area		**0.835**		**0.835**		**0.835**		0.815		**0.835**		**0.835**		**0.816**	
Attributes	1	Geology		Geology		Geology		Geology		Geology		Geology		Geology	
	2	GeoMorph		GeoMorph		GeoMorph		GeoMorph		GeoMorph		GeoMorph		GeoMorph	
	3	Streams		~~Streams~~		~~Streams~~		~~Streams~~		~~Streams~~		~~Streams~~		~~Streams~~	
	4	Faults		Faults		Faults		~~Faults~~		Faults		Faults		Faults	
	5	Land use		Land use		~~Land use~~		~~Land use~~		~~Land use~~		~~Land use~~		~~Land use~~	
	6	Slope		Slope		Slope		Slope		Slope		Slope		Slope	
	7	Aspect		~~Aspect~~		~~Aspect~~		~~Aspect~~		~~Aspect~~		~~Aspect~~		~~Aspect~~	
	8	TWI		TWI		TWI		TWI		~~TWI~~		~~TWI~~		~~TWI~~	
	9	Curvature		Curvature		Curvature		Curvature		Curvature		~~Curvature~~		~~Curvature~~	
	10	Plan Curv.		~~Plan Curv.~~		~~Plan Curv.~~		~~Plan Curv.~~		~~Plan Curv.~~		~~Plan Curv.~~		~~Plan Curv.~~	
	11	Prof. Curv.		Prof. Curv.		Prof. Curv.		Prof. Curv.		Prof. Curv.		Prof. Curv.		~~Prof. Curv.~~	

Model Bay-5 has the highest F-measure of LS points, 0.621, and second highest for NLS points, 0.834 (Table 6.4). Model Bay-6 has the second highest F-measure of LS points, 0.619, and the highest for NLS points, 0.839 (Table 6.4). Besides recall of LS points and precision of NLS points, models Bay-5 and Bay-6 have the majority of all the superior results, with Bay-6 having the most optimized results for the different measures (Table 6.4). Bay-7 has the highest recall of LS points, 0.848, but that model also had the lowest recall of NLS points, 0.602, out of all the models (Table 6.4). The same is true in reverse for the NLS and LS points with precision in model Bay-7 (Table 6.4).

Choosing the superior model is not straight forward with Bayesian Networks, unlike with J48 and logistic, the best results of the different measures for Bayesian Networks are not exclusively found in one model. While Bay-6 has the majority of the highest optimized measures, Bay-5 has a superior overall accuracy, 0.761 compared to 0.757, but since the difference is 0.4%, it's negligible, same is true with the F-measure of LS (Table 6.4). This study is trying to predict where LS will occur with the most accuracy, whereby these models are attempting to optimize F-measure, recall and precision of LS points without overly degrading NLS points, which in all algorithm iterations has remained very high (Table 6.2, Table 6.3 and Table 6.4). Bay-6 has neither the highest or second highest recall of LS, it does have the highest LS F-measure and precision score. While Bay-5 does have a higher recall of LS points, Bay-6 has the optimized recall of NLS points, in other words, the least amount degradation in this variable for NLS points while optimizing the models ability to predict LS points.

Table 6.5 compares the optimized J48, logistic and BayesNet model. Model J48-3 eliminated only aspect and plan curvature, attributes that the all attribute analysis reported to have little value (Table 6.1). Models Log-7 and Bay-6 used the same attribute combination: geology, geomorphology, faults, slope and profile curvature (Table 6.5). The attribute evaluators IG, GS and CS had geology, geomorphology, slope and profile curvature in the top four rankings, but ranked distance to faults as 7 on average (Table 6.1). Attribute evaluator FS predicted geomorphology, slope and profile curvature to have 10 folds, distance to faults had 2 and geology 1 fold (Table 6.1). FS indicated all the other attributes would have no influence, which proved true in models Log-7 and Bay-6 (Table 6.1 and Table 6.5).

To choose the best model from Table 6.5 the statistical parameters had to be compared between the three ideal models: J48-3, Log-7 and Bay-6. Of the three models, Log-7 has superior results for F-measure of NLS points, recall of NLS, precision of LS, correctly identified instances and ROC area (Table 6.5). Log-7 also has the lowest results for F-measure of LS points, recall of LS points, precision of NLS points and overall accuracy (Table 6.5). Model Bay-6 has one of the highest recalls of LS points without severely degrading the recall of the NLS (Table 6.5). Bay-6 also has one of the best F-measures for LS points and overall accuracy (Table 6.5).

The best model for landslide susceptibility will optimize the measures for LS points, while minimizing degradation of the NLS parameters. As model iterations (Table 6.2, Table 6.3 and Table 6.4) showed improvement for LS parameters sometimes the NLS measures degraded insignificantly, but in other instances they degraded severely. The LS measures needed to be optimized to increase the confidence the landslide probability predictions, but since the NLS measures reached over 0.800 and higher in some of the models it’s also important to maintain a similar level of confidence.

Table 6.5: The optimized models for each algorithm. The highest number of each parameter, across all models, has red text and the second highest parameter is shown in bold. The best model is outlined in blue.

Top J48, Log and Bay Models		**J48-3**		**Log-7**		**Bay-6**	
		KNLS	KLS	KNLS	KLS	KNLS	KLS
Confusion Matrix	PNLS	499	61	513	47	443	117
	PLS	90	101	101	90	53	138
F-Measure of LS		**0.572**		0.549		**0.619**	
F-Measure of NLS		**0.869**		**0.874**		0.839	
Recall of LS		**0.529**		0.471		**0.723**	
Recall of NLS		**0.891**		**0.916**		0.791	
Precision of LS		0.623		0.657		0.541	
Precision of NLS		0.847		0.836		0.893	
Correctly Identified*		**79.89%**		**80.29%**		77.36%	
Overall Accuracy		**0.710**		0.694		**0.757**	
ROC Area		0.783		**0.846**		**0.835**	
Attributes	1	Geology		Geology		Geology	
	2	GeoMorph		GeoMorph		GeoMorph	
	3	Streams		~~Streams~~		~~Streams~~	
	4	Faults		Faults		Faults	
	5	Land use		~~Land use~~		~~Land use~~	
	6	Slope		Slope		Slope	
	7	~~Aspect~~		~~Aspect~~		~~Aspect~~	
	8	TWI		~~TWI~~		~~TWI~~	
	9	Curvature		~~Curvature~~		~~Curvature~~	
	10	~~Plan Curv.~~		~~Plan Curv.~~		~~Plan Curv.~~	
	11	Prof. Curv.		Prof. Curv.		Prof. Curv.	

Model Bay-6 was chosen to build the landslide probability map, due to the optimized model statistical measures for both LS and NLS points (Figure 6.1 and Table 6.5). Bay-6 included the attributes geology, geomorphology, distance to faults, slope and profile curvature. One category of the geology and geomorphology each follow the arc of the caldera wall, typifying the steepest slopes in the study area (Figure 4.5, Figure 4.6, amd Figure 4.10). Faulting in a region can affect the slopes, meaning the faults could be an indirect way of looking at slope angles (Figure 4.8 and Figure 4.10). And profile curvature is a derivative of the slope angles of the study area (Figure 4.10 and Figure 4.13). Since all the attributes spatial distribution could have a direct relationship to slope, thus weighting the models by using layers potentially favoring slope, all three algorithms were computed with just the layer slope. The results are show in with just the confusion matrices in Table 6.6. While the NLS were predicted with a lot of success, all three algorithms had a zero or near zero recall for LS points (Table 6.6).

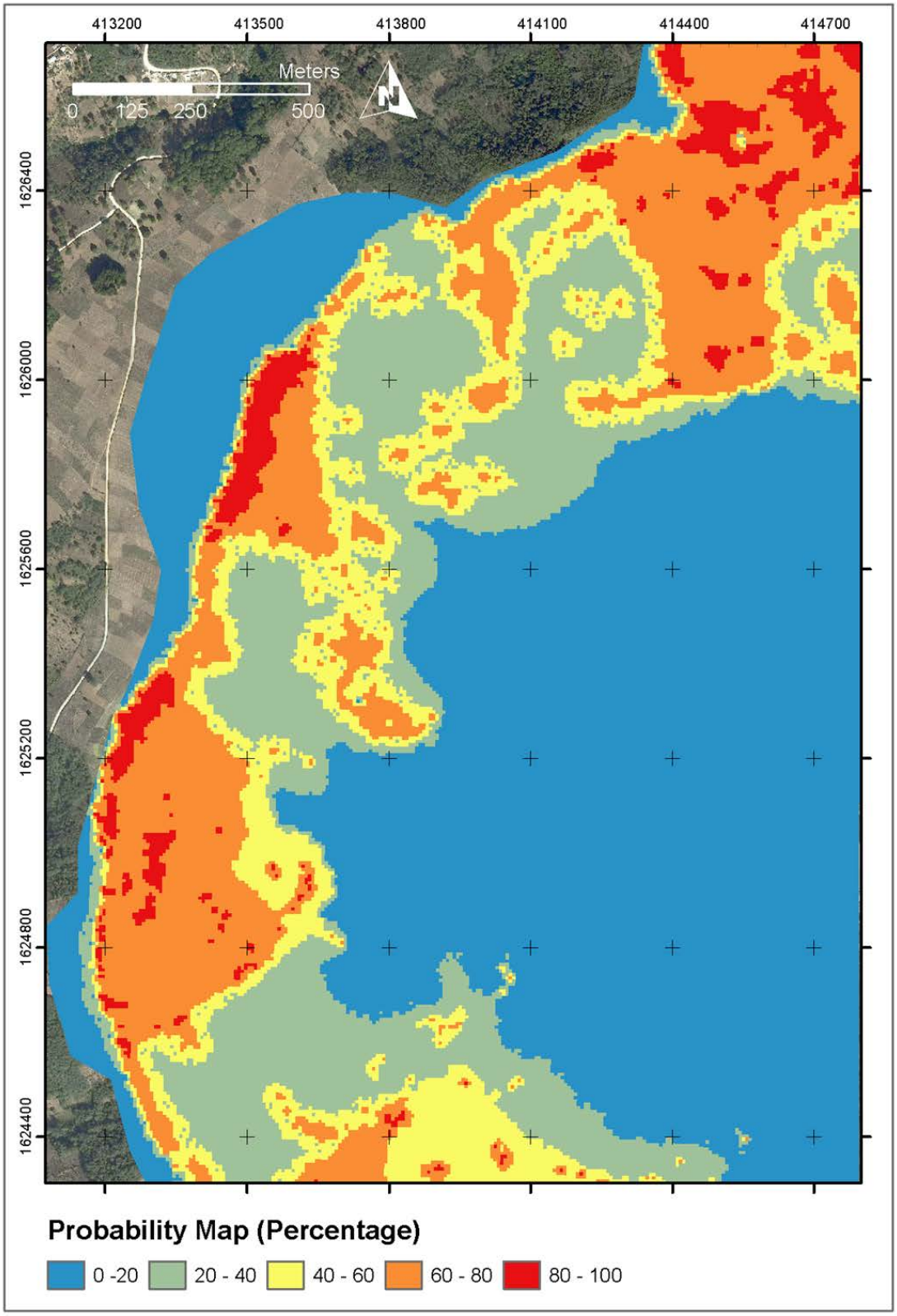


Figure 6.1: Landslide Probability Map of San Juan. See Table Table 6.7 for more information

Table 6.6: The algorithms computed with just the slope attribute.

Algorithms with just slope attribute		J48-3		Log-7		Bay-6	
Confusion Matrix		KNLS	KLS	KNLS	KLS	KNLS	KLS
	PNLS	529	31	517	43	560	0
	PLS	170	21	159	32	191	0

To build the probability map Model Bay-6 was chosen, because it optimized the statistical measures for both LS and NLS points (Table 6.5). Figure 6.1 shows the probability map of the study area developed from data set C (the 10,000 random points) processed through model Bay-6, which used the attributes: geology, geomorphology, faults, slope and profile curvature. The probabilities were divided into five classes: 0-20%, 20-40%, 40-60%, 60-80% and 80-100% (Figure 6.1 and Table 6.7). Five percentage classes were chosen so five hazard classes could be assigned: very low, low, moderate, high and very high (Table 6.7). Five classes were chosen so the result of this multivariate approach could be compared to the results of the GM bivariate approach (2012a) that used 5 hazard classes.

Table 6.7: Calculations of landslides, study area coverage and landslides per hazard class area for the probability map of this study using the LS points from data set B. Note that the total number of landslidse is less than data set B. This discrepancy resulted from the holes created in the hazard zones when the rasters were converted to polygons in ArcMap.

Probability (%)	Hazard Class	# of LS	% of total LS	Area of Hazard Class (km)	% of Study Area	Landslide Density per Area of Hazard Class (km)
0 – 20	Very Low	2	1	2405.3	57.03	0.018
20- 40	Low	31	17	647.9	15.36	1.11
40 - 60	Moderate	31	17	442.6	10.49	1.62
60 – 80	High	74	40.4	610.1	14.46	2.79
80 – 100	Very High	45	24.6	111.9	2.65	9.28
Total		183	100	4217.8	99.99	

Table 6.8: Calculations of landslides, study area coverage and landslides per hazard class area for the GM (2012a) map of this study using the LS points from data set B.

Probability (%)	Hazard Class	# of LS	% of total LS	Area of Hazard Class (km)	% of Study Area	Landslide Density per Area of Hazard Class (km)
0 – 20	Very Low	0	0.0	916.0	26.9	0.11
20- 40	Low	13	6.8	770.9	22.7	0.17
40 - 60	Moderate	24	12.6	272.7	8.0	0.90
60 – 80	High	129	67.5	1134.9	33.3	2.25
80 – 100	Very High	25	13.1	308.7	9.1	3.29
	Total	191	100	3403.2	100	

To evaluate the success of the map created by this study as compared to the GM map (2012a), data set B was overlain each map (Figure 6.2 and Figure 6.3). Then the number of landslides per class was counted and the percentage of landslides each class captured was calculated (Table 6.7 and Table 6.8). The general pattern for number of LS found in each hazard class for both the multivariate and bivariate maps are an increase from very low to high, with a decrease to very high (Table 6.7 and Table 6.8). For the area of the hazard class for both maps, the area decreases from very low to moderate, increases at high and decreases again at very high (Table 6.7 and Table 6.8). The more successful approach will capture a higher percentage of LS in a smaller area, which can be seen in the landslide density per area of hazard class column in Table 6.7 and Table 6.8. The relationship of landslide density per area of hazard class can be seen in (Figure 6.4).

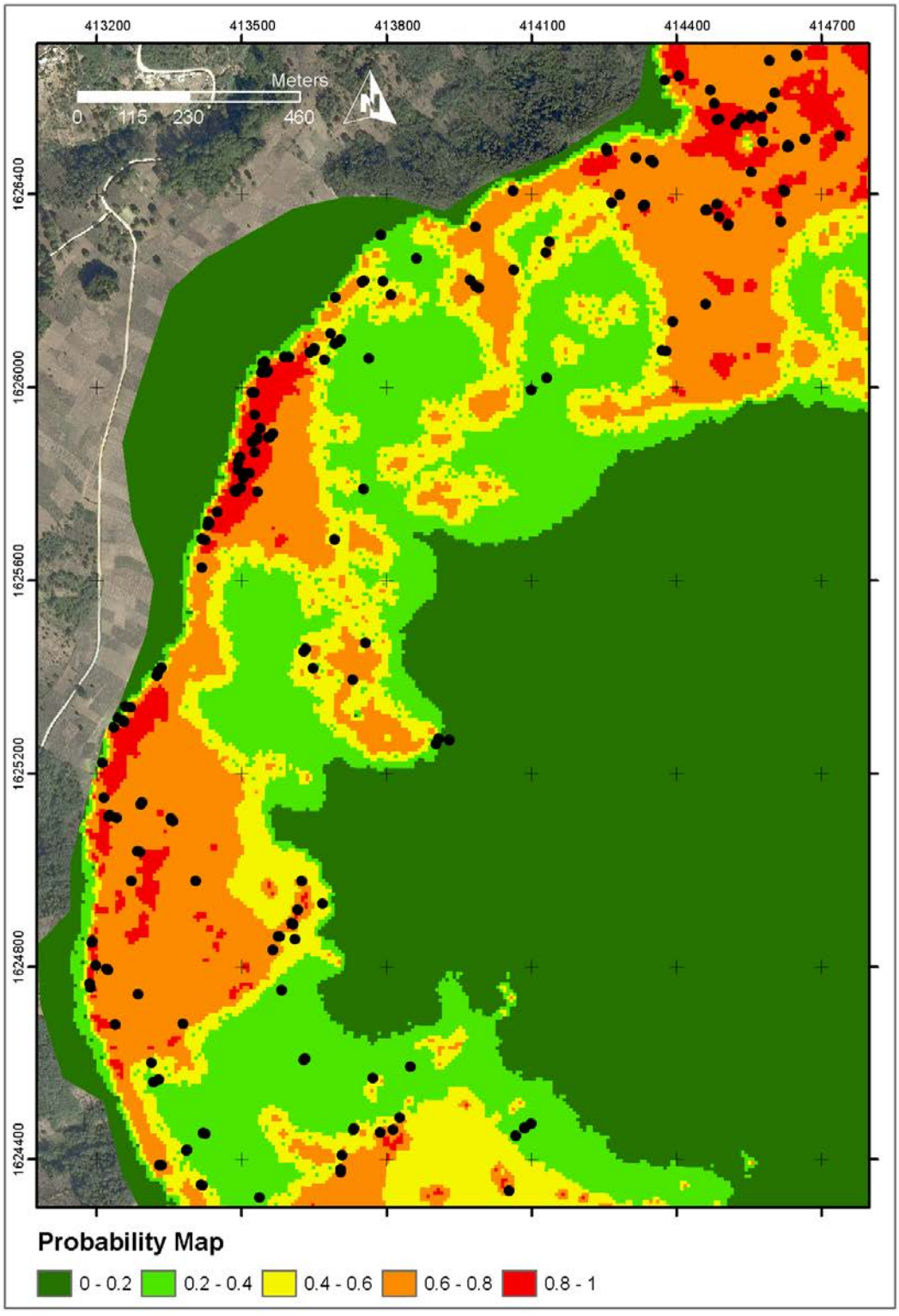


Figure 6.2: The probability map using data set C and model Bay-6. The LS points used to develop model Bay-6 have been overlain.

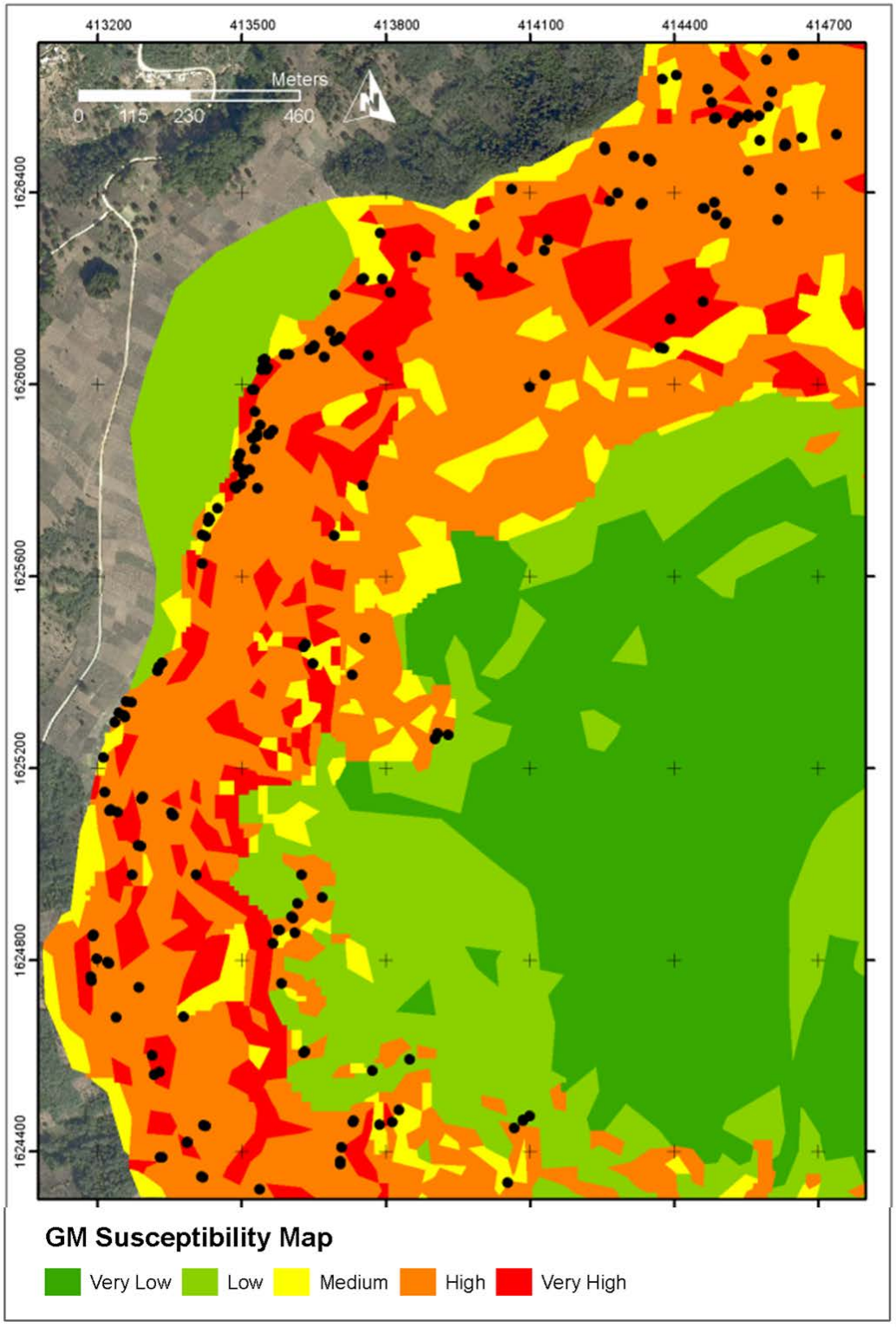


Figure 6.3: The GM map. The LS points used to develop model Bay-6 have been overlain.

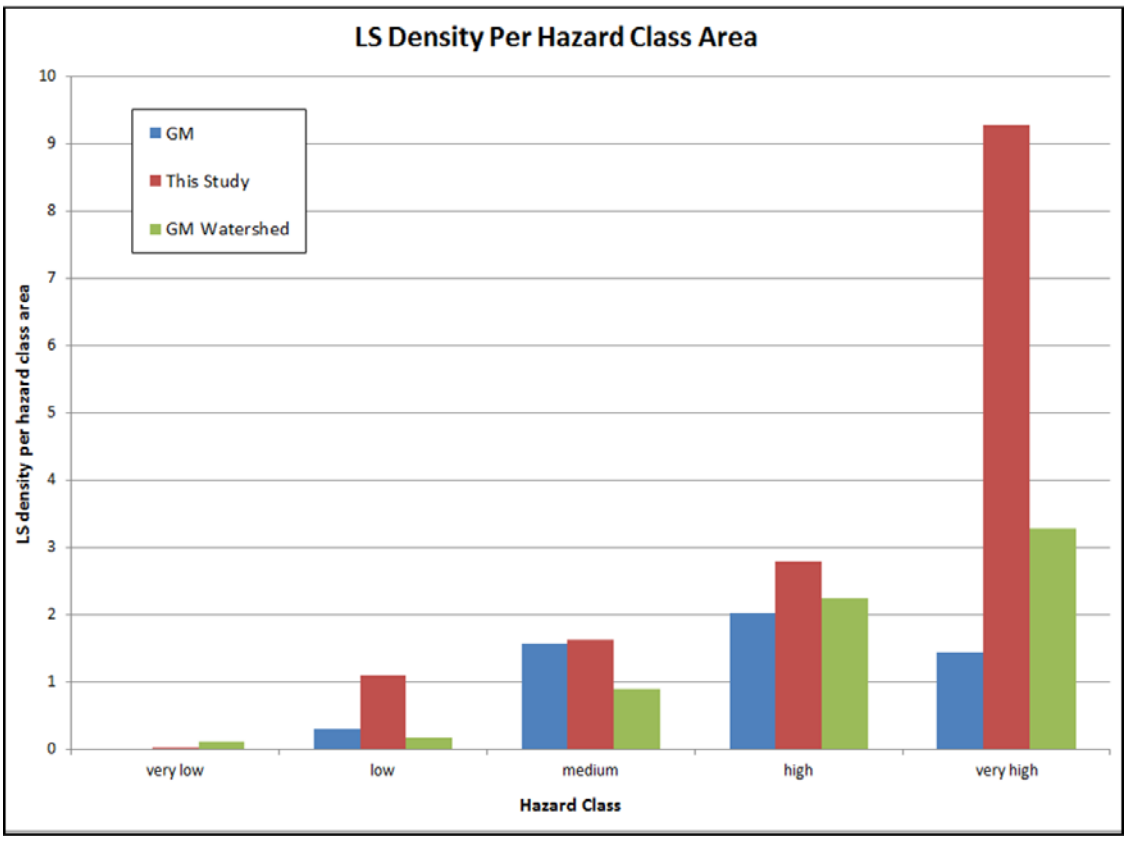


Figure 6.4: Landslide density per hazard class area. This study and the GM (2012a) map used data set B, and the GM watershed used the complete Hurricane Stan landslide database.

Figure 6.4 shows the landslide density per hazard class area increasing from very low to high, with a decrease at very high for the GM map that covered the study area used in this study. The landslide density per hazard class density for this study showed an increase from very low to very high, with three-fold increase from classes high to very high (Figure 6.4). A comparison between this study's probability map and the portion of GM map that intersects with this study area would show that the multivariate approach is more successful at capturing more landslides in a smaller area the higher the hazard class becomes (Figure 6.4). To better understand this relationship it is important to consider that the GM map (2012a) used the entire database of landslides from Hurricane Stan over the entire watershed of Lake Atitlán.

To better evaluate the accuracy of the GM (2012a) model the same calculations found in Table 6.7 and Table 6.8 were completed for the whole GM map of the watershed. The results for landslide density per hazard class are shown in Table 6.9 and

Figure 6.4. Unlike the portion of the GM map that overlapped the study area that showed less accuracy in capturing more landslides per area in the very high class, the results for the whole watershed show a steady increase from hazard class very low to very high (Figure 6.4). Unlike the 3-fold increase observed in the multivariate analysis from high to very high, the result for the whole watershed only showed an increase of a third from high to very high (Figure 6.4).

Table 6.9: Calculations of landslides, study area coverage and landslides per hazard class area for GM (2012a) map of the watershed using all the Hurricane Stan Landslides.

Probability (%)	**Hazard Class**	**# of LS**	**% of total LS**	**Area of Hazard Class (km)**	**% of Study Area**	**Landslide Density per Area of Hazard Class (km)**
0 – 20	Very Low	161	0.9	33,615,432	8.0	0.11
20- 40	Low	1133	6.2	154,229,621	36.5	0.17
40 - 60	Moderate	4442	24.3	114,555,438	27.1	0.90
60 – 80	High	9821	53.7	100,870,736	23.9	2.25
80 – 100	Very High	2738	15.0	19,243,111	4.6	3.29
	Total	18295	100	422,514,337	100	

The increase in number of landslides captured as the hazard class area decreases for the multivariate approach and the bivariate approach for the whole watershed (Figure 6.4) is a typical relationship found in other similar studies (Suzen et al. 2004; Ozdemir 2011; Song et al). This relationship does not hold true for the portion of the GM map (2012a) that overlaps with the study area around San Juan (Figure 6.4). Indicating that at a smaller scale the multivariate approach is more accurate, and at a larger scale the GM bivariate approach is also successful. At a smaller scale the GM (2012a) bivariate approach is less accurate for this study area. Other investigations across the watershed for the GM map (2012a) were not taken to evaluate success at smaller scales by looking at landslide density per hazard class area.

Figure 6.5 shows the landslide probability map using model Bay-6 validated with landslides resulting from Tropical Storm Agatha. Of the seven different landslide from Agatha that overlap with the study area, three originate from the area with the highest risk (80 – 100%), one originates from the second highest risk zone (60 – 80%), two initiated in the moderate risk zone (40 – 60%), with the last one initiating in the second lowest risk (20 – 40%) as seen in Figure 6.5.

Figure 6.6 overlays the landslides from Agatha over the GM (2012a) susceptibility map to independently validate their results, as is seen in Figure 6.5 for the probability map produced by this study. Of the seven landslides from Agatha, 1 is on the border between the high and very high class, 5 fall in the high hazard zone and 1 falls within the very high hazard zone bounds (Figure 6.6). The multivariate probability map had 3 Agatha landslides fall in the top hazard rank as compared to the 1 in the GM map (2012a). The GM had all the Agatha slides initiating from the top two ranks, compared to the multivariate probability map shows 6 of the 7 slides initiating from the top three ranks.

It's possible that the GM map (2012a) is more conservative in their predictions, seeing that all the landslides initiated from the top two ranks. Even though the Agatha landslides initiated from more of the hazard classes on the probability map created by this study, only one initiated in the 20 – 40%, with majority of the landslides initiating in the top two ranks. Using the Agatha landslides to independently validate the multivariate and bivariate (GM 2012a) approaches shows them to be comparable in that in both maps the landslides are initiating from the higher hazard classes (Figure 6.5 and Figure 6.6). But 5 of the 7 Agatha landslides initiate from the high class with one the border, which again displays the relationship that the bivariate method captures more landslides in the in high class than the very high class (Figure 6.4). The relationship observed in Figure 6.4 that the multivariate approach captured more landslides for the smaller areas of the higher hazard classes. This relationship is displayed in Figure 6.5 as far as that 3 Agatha slides initiated from the high class, the class with the smallest area, with the number of slides captured per class between very low and high being 1 slide or less. Since only 7 Agatha slides fell within the study area is hard to truly see if the relationships observed in Figure 6.4 is maintained.

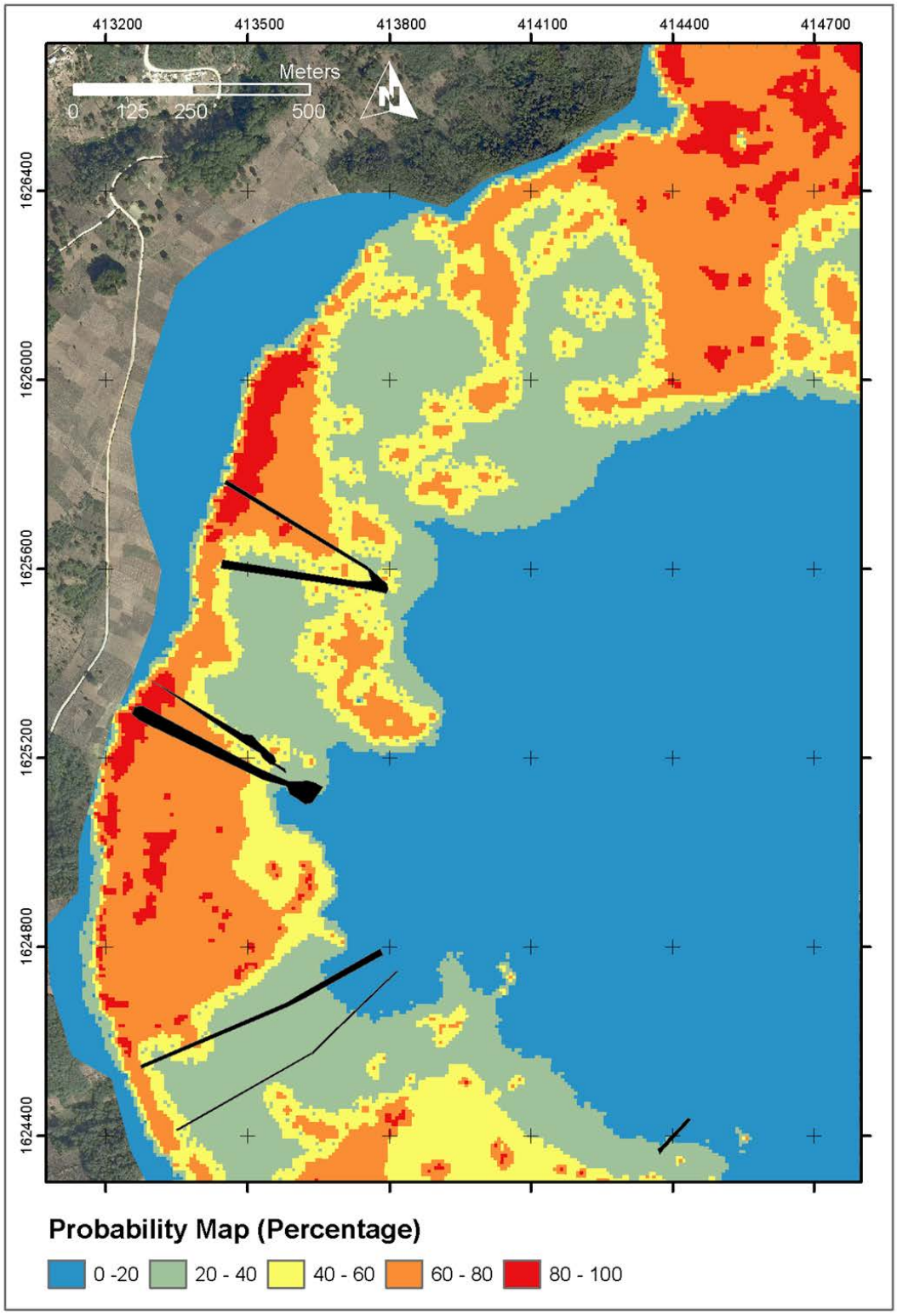


Figure 6.5: Landslide Probability Map Validated with Agatha Landslides. See Table Table 6.7 for more information.

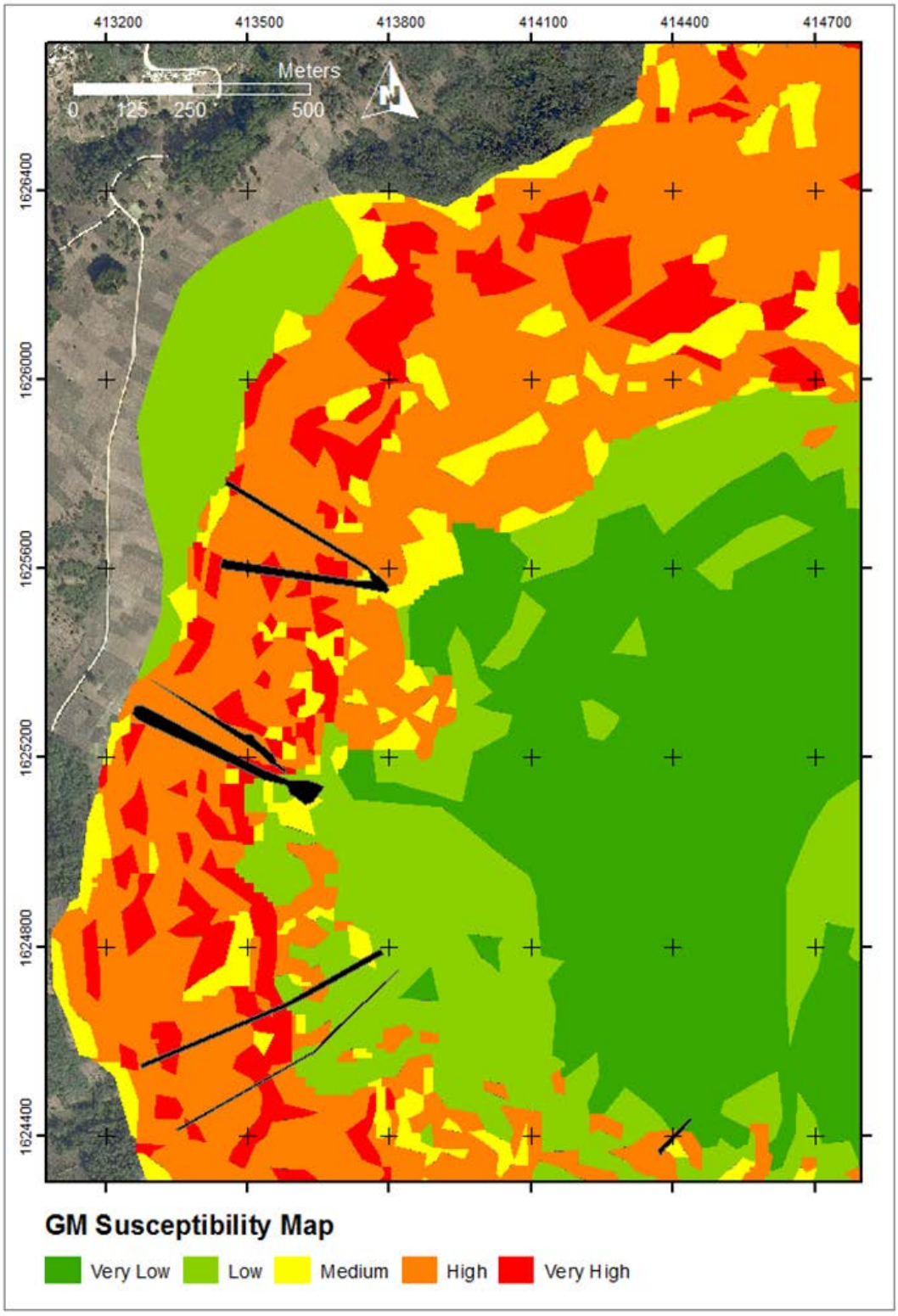


Figure 6.6: GM Hazard Map with Agatha Landslides overlain.

The probability map created with Bay-6 is based off data collected shortly after hurricane Stan. The probability map is only applicable to rain events comparable to Stan. Comparative events can be seen by looking at yearly precipitation data for the region. Similar cumulative rain patterns that led to the Stan landslides can also be seen for the 2010 Agatha precipitation data and the 2011 Tropical Storm 12e data (Figure 6.7).

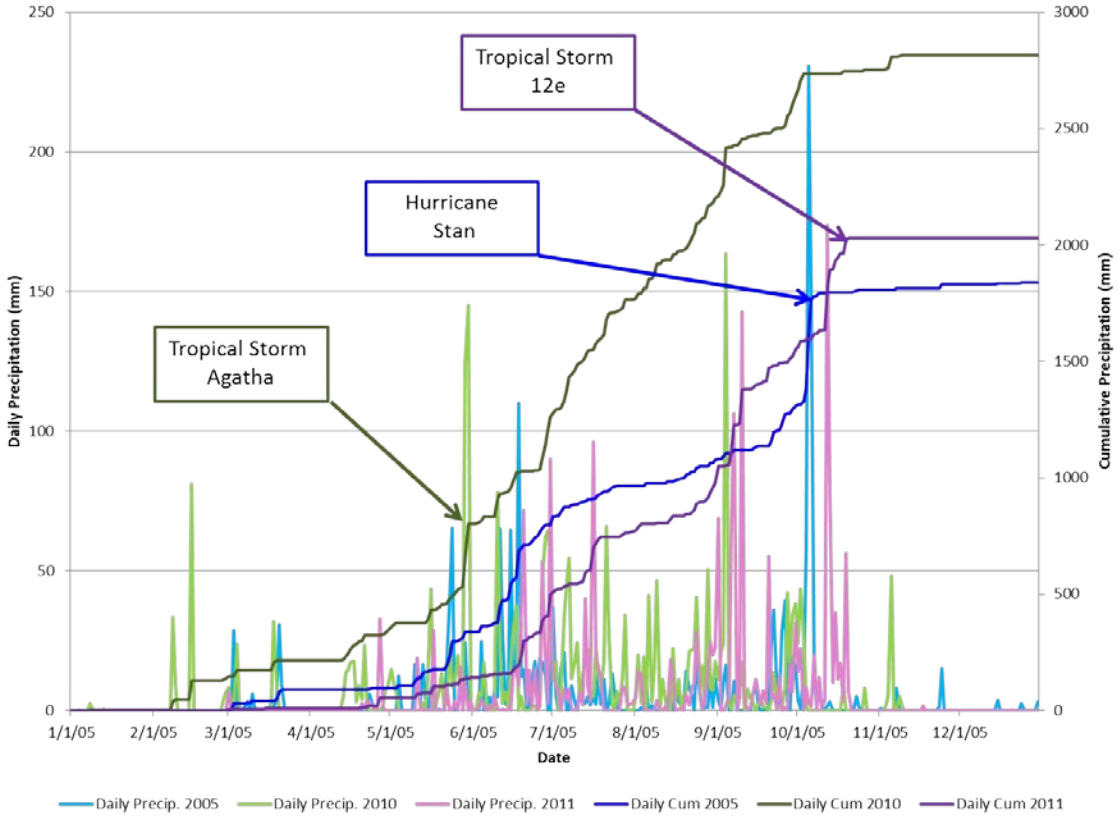


Figure 6.7: Graph of Precipitation for years 2005, 2010, and 2011.

The peak rainfall during Hurricane Stan occurred in early October; for Agatha early of June; and, again, early October for Storm 12e (Figure 6.7). While landslides did occur with storm 12e, there were much fewer than with Agatha and Stan, and they happened outside the bounds of the study area.

Data set B was used to train the model and run data set C to create the spatially interpolated susceptibility map. The map was independently verified with Agatha landslides, a storm that occurred almost five years after the data that was used in this study. Data set B was not initially validated, point by point, to ensure LS and NLS were properly identified; Data set A, its predecessor, relied the unverified 200 LS picks and on the randomly chosen NLS points, using ArcTool *Spatially Balance Point.* Neither the LS or NLS points in data set A were verified using the orthophoto as was used for data set B. In a small study area, such as the case here, it was feasible to verify, point by point, that each point chosen was correctly identified, creating a very accurate data set B. On a larger scale, like the watershed of Lake Atitlán, hand verifying almost 20,000 randomly chosen LS points plus the tens of thousands of NLS points required to model would be time prohibitive. Though the model for the watershed could be run using the arbitrary centroids on the landslide scarps as identified by GM (2012a) in the Hurricane Stan landslide dataset, confirming that all the *Spatially Balance Points* chosen to be NLS were actually NLS would still be time prohibitive. And since the idea is to make this procedure easily and speedily reproducible, having to confirm tens of thousands of points every time a hazard map is to be updated, it is not efficient. To begin to quantify how usable data set A would be in creating a susceptibility map, the data set was run through model Bay-6 to compare its results to dataset B (Table 6.10).

While every statistical measure is lower for dataset A (Table 6.10), only 8 less LS points were correctly identified, as observed in the confusion matrix, and most measures decrease by less than 10%. Further, the recall of the LS points for data set A is superior to the results of dataset B in models J48-4 and Log-7 (Table 6.5 and Table 6.10), an important difference due to these models successes being based on the optimization of the recall of the LS points.

Table 6.10: Bayesian Network Algorithm Results for Dataset A and B

BayesNet (Bay-6)		Dataset B		Dataset A	
		KNLS	KLS	KNLS	KLS
Confusion Matrix	PNLS	443	117	508	143
	PLS	53	138	70	130
F-Measure of LS		0.619		0.550	
F-Measure of NLS		0.839		0.827	
Recall of LS		0.723		0.650	
Recall of NLS		0.791		0.780	
Precision of LS		0.541		0.476	
Precision of NLS		0.893		0.879	
Correctly Identified*		77.36%		74.97%	
Overall Accuracy		0.757		0.715	
ROC Area		0.835		0.806	
Attributes	1	Geology		Geology	
	2	GeoMorph		GeoMorph	
	3	~~Streams~~		~~Streams~~	
	4	Faults		Faults	
	5	~~Land use~~		~~Land use~~	
	6	Slope		Slope	
	7	~~Aspect~~		~~Aspect~~	
	8	~~TWI~~		~~TWI~~	
	9	~~Curvature~~		~~Curvature~~	
	10	~~Plan Curv.~~		~~Plan Curv.~~	
	11	Prof. Curv.		Prof. Curv.	

7.0 Conclusions

The data used to create the probability map of San Juan La Laguna was developed from the data points of LS and NLS points found in data set B and the information from the eleven primary and secondary attributes. To determine which of the attributes were more likely influencing the occurrence of landslides four attribute analyzers in Weka (Hall et al. 2009) were used to rank the data: IG, GR, CS, and FS. Next the algorithms that yielded the best preliminary results were identified using classifiers: decision tree J48, logistic regression and BayesNet. Using the rankings from the attribute analyzers as guide, each of the 3 models were optimized for LS prediction. The optimized models of each algorithm were compared, showing Bay-6 to be the best model. Also, the results of Bay-6 from data set B was compared to the results obtained using data set A. Data set A consisted of unverified LS and NLS points that had been randomly chosen to be spatially balanced. Data set A was an attempt to quantify the error that could exist when a data set became too large to efficiently verify point by point. The probability map was then created through spatial interpolation, using the Bay-6 model from data set B, and

independently validated with the landslides that resulted from Tropical Storm Agatha in 2010. Finally, the success and methodology of the multivariate model used here was compared to the bivariate method used by GM (2012a) through the comparison of landslides percentage in each zone and the independent validation of the Agatha landslides.

The four attribute analyzers, IG, GR, CS and FS, proved useful in determining which attributes should be eliminated, especially in the first one to three model iterations (Table 6.1). In seeking to optimize the LS measures, the rankings from attribute analyzers IG, GR and CS held less value, what they suggested as the top layers to keep in the model did not contribute to optimized models for each of the algorithms. The final attribute combination for algorithm J48 eliminated only slope aspect, plan curvature and profile curvature (Table 6.2). Removing any other attributes contributed to model degradation (Table 6.2). Attribute analyzer FS proved very accurate in determining the ideal attribute combination for logistic regression and BayesNet, saying only attributes geology, geomorphology, slope, distance to faults and profile curvature would be landslide indicative (Table 6.1, Table 6.2, Table 6.3, Table 6.4, and Table 6.5).

Of the three algorithms that showed the most initial success (decision tree J48, logistic regression and BayesNet), BayesNet was the best at predicting LS points while maintaining the credibility of the NLS predictions (Table 6.5). Bay-6 had the highest LS F-measure, LS recall and overall accuracy of the three models (Table 6.5).

Comparing the Bay-6 model results for data set A and B showed that verifying a LS and NLS point where correctly identified improved the success of the model (Table 6.10). The study area used here was small enough making the point-by-point validation feasible, but this study also aimed to make this method easily reproducible. This study area falls within a larger watershed that experiences the same landslide hazard. Ideally this model would be applied to the whole watershed, which experienced almost 20,000 landslides because of Hurricane Stan in 2005. Data set B showed that in smaller areas this method could be easily applied to other parts of the Lake Atitlán watershed. Data set A showed the model was still reasonably successful even without point by point validation. It would just be importation to remember as information was used from maps developed

in a method similar to data set A, that the data be used conservatively for hazard mitigation and planning.

The compilation of the number of landslides per division of probability and the area each class covered for this study is shown in Table 6.7. While the transition from the "high" class to the "very high" class did not follow the traditional pattern of continued increase in number of landslide and decrease in area covered (Suzen et al. 2004; Ozdemir 2011; Song et al), this study chose to use 5 divisions where most studies choose four. Five divisions were chosen to better compare the results of this probability map, using a multivariate approach, to the bivariate approach used in the GM (2012a) study. Further, considering the more than 10% decrease in area between the two highest hazard classes for this study (Table 6.7) compared to the typical change of only a few percentage points (Suzen et al. 2004; Ozdemir 2011; Song et al), the decrease in number of landslides is explainable.

Comparing the density of landslides per hazard class area of this study and the GM study showed that on a small scale the bivariate method was less accurate in capturing a higher ratio of landslides to a smaller area in the highest hazard class (Table 6.7, Table 6.8, and Figure 6.4). Further investigation of the bivariate study looking at the whole watershed and all the landslides used to develop their susceptibility map showed the results to be more accurate at a larger scale (Figure 6.4 and Table 6.9).

The independent validation of the probability map using the landslides from Tropical Storm Agatha showed the success with which Bay-6 predicted the occurrence of landslides. Four of the seven landslides that occurred in the study area, because of Agatha, fell within the top two probability divisions, with three of the four initiating in the top probability zone (Figure 6.5). Both the probability map created here and the GM map (2012a) had the most at risk slopes following the steep crater walls. Where only four of seven landslides fell in the top to hazard zones in the results of this study, the GM map (2012a) showed all the 7 Agatha landslides initiating from the top two hazard zones, which could just a conservative estimate of the most at risk slopes.

The probability map created from Model Bay-6 and the GM map (2012a) are fairly comparable. While a direct quantitative approach, like comparing the ROC area as

Yesilnacar et al. (2005) demonstrated, was not possible, comparing the independent validations using the Agatha landslides proved informative. The advantage to the Bay-6 model is the ideal combination of attributes have been identified subjectively.

8.0 Limitations and Future Work

The probability map created by this study utilized a very small portion of the available landslide data set in the Lake Atitlán watershed. A probability map for the whole watershed could be developed using this methodology in one of two ways: one, select random points from the whole watershed and run them in model Bay-6, and two, use the watershed landslide data base from Hurricane Stan, select NLS points and rerun attribute combinations in Weka. The first option may bias the probability map of the watershed to conditions specific to the study area surrounding San Juan, but this option would be relatively faster. Option two acknowledges that BayesNet predicts LS points more reliably, but also the need to identify how attributes behave across the watershed. As van Westen (1997) highlighted, not every landslide in a study area necessarily experiences the attributes in the same way, causing failure. A way to compensate for this assumption is to run the multivariate analysis on restricted area sizes of the watershed, creating numerous subsets of the landslide database from Hurricane Stan.

The most time consuming part of option two would be creating the spatially balanced NLS points. After the data set of LS, NLS and attributes were created and imported into Weka, the ideal combination of attributes for the watershed could be identified. Once the optimized model was identified it should be compared to the result of Bay-6 in this study, remembering the slight discrepancy in unverified points as seen in data set A (Table 6.10). Next, random points across the watershed should be chosen to create a data set of all unknown points to run through the optimized watershed model, using the results to interpolate the probability map. If ArcMap was not available the program R (R Core Team 2013), open source software, is capable of dealing with rasters and could be used to create the probability map.

The probability map was created using the data collected right after Stan occurred in early 2006 and the randomly selected points within that data set. Seven rainy seasons

have passed, the most notable being Agatha and 12e. While the map was validated using the landslides of Agatha, the model would be further improved by acquiring a new DEM to update the attributes and rerun the model using Bay-6. After the new map was created with the updated information, it could be compared to the probability map created here; and likely, the two probability maps could be fairly similar. The probability map based of the new DEM may highlight new areas that were not high risk areas found in this study, as surficial deposits change due to erosion and sliding. One known attribute that is not yet complete is a water table map of the watershed (GM 2012a), which should soon be made available, thus, easily incorporated into the methods posed here.

One limitation with this dataset was an accurate representation of how water flowed on the surface and could infiltrate the ground, affecting the soil. TWI was used to investigate how runoff could affect the model, but it assumed steady state conditions. If soil properties could be established, cohesion parameters could be incorporated into the model, potentially increasing the accuracy of the model. Also, the probability map focuses on areas most likely to experience failure, it does not model how different volumes of material would behave in different failure scenarios. Modeling flow hazards would further help the community of San Juan.

The probability map can be used by the town of San Juan for future land use planning. Further, the municipality can hold workshops teaching the people where to expect landslides to occur during high-yield rain events; many of the townspeople work in fields on those slopes or at the bottom of where the landslides would inevitably flow through.

References

Aleotti, P., & Chowdhury, R. (1999). Landslide hazard assessment: summary review and new perspectives. *Bulletin of Engineering Geology and the Environment*, *58*(1), 21-44.

Bharti, K., Jain, S., & Shukla, S. (2010). Fuzzy K-mean Clustering Via J48 For Intrusiion Detection System. *Kusum Bharti International Journal of Computer Science and Information Technologies (IJCSIT)*, *1*(4), 315-318.

Chen, Z., & Wang, J. (2007). Landslide hazard mapping using logistic regression model in Mackenzie Valley, Canada. *Natural Hazards*, *42*(1), 75-89.

Chung, C. J. F., & Fabbri, A. G. (1999). Probabilistic prediction models for landslide hazard mapping. *Photogrammetric Engineering and Remote Sensing*,*65*(12), 1389-1399.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and \ psychological measurement*, *20*(1), 37-46.

CONAP (Consejo Nacional de Áreas Protegidas). (2007); Plan Maestro de la Reserva de Uso Múltiple Cuenca del Lago Atitlán: 2007-2011. Gobierno de Guatemala.

ESRI (Environmental Systems Resource Institute): ArcTool Create Spatially Balanced Points. 2012a. ArcMap 10.1. ESRI, Redlands, California. http://resources.arcgis.com/en/help/main/10.1/index.html#//00310000009z000000

ESRI (Environmental Systems Resource Institute): How IDW Works. 2012b. ArcMap 10.1. ESRI, Redlands, California. http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/How_IDW_works/009z00000075000000/

ESRI (Environmental Systems Resource Institute): Understanding Curvature Rasters. 2012b. ArcMap 10.1. ESRI, Redlands, California. http://blogs.esri.com/esri/arcgis/2010/10/27/understanding-curvature-rasters/

Fernandez-Lavado, C. (2008); Manual metodologico para la evaluacion de movimientos de ladera en el AMSS (El Salvador, CA). Programa IPGARAMSS. Finance by European Union, AECID, Ayuntamiento de Barcelona, Area Metropolitana de Barcelona, Diputacion de Barcelona. Executed by OPAMSS, COAMSS and Geologos del Mundo. 54 pgs.

GM (Geólogos del Mundo). (2012a); Guía metodológica para la elaboración de mapas de susceptibilidad: A movimientos de ladera en la cuenca del lago atitlán, Guatemala. *Con el apoyo financier de la Agencia Española de Cooperación Internacional para el Desarrollo (AECID).*

GM (Geólogos del Mundo). (2012b); Mapa geológico de la cuenca del Lago de Atitlán 1:25.000. Guatemala.

GM (Geólogos del Mundo). (2012c); Mapa geomorfológico de la cuenca del Lago de Atitlán 1:25.000. Guatemala.

Hall, Mark, Frank, Eibe, Holmes, Geoffrey, Pfahringer, Bernhard, Reutemann, Peter, and Witten, Ian H. (2009); The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1.

Hoehler, F. K. (2000). Bias and prevalence effects on kappa viewed in terms of sensitivity and specificity. *Journal of clinical epidemiology*, *53*(5), 499-503.

Hwang, S., Guevarra, I. F., & Yu, B. (2009). Slope failure prediction using a decision tree: A case of engineered slopes in South Korea. *Engineering Geology*, *104*(1), 126-134.

Keyport, R. (2013). Comparison of pixel-based versus object-oriented classification of landslides using remote sensing data. (Unpublished MS report, Michigan Technological University).

Lee, S., & Min, K. (2001); Statistical analysis of landslide susceptibility at Yongin, Korea. *Environmental Geology*, *40*(9), 1095-1113.

Lee, S., & Pradhan, B. (2007). Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models. *Landslides*, *4*(1), 33-41.

Luna, B. Q. (2007). Assessment and modelling of two lahars caused by" Hurricane Stan" at Atitlan, Guatemala, October 2005. University of Oslo.

MAGA. (2011); Mapa de Usos del Suelo del departamento de Sololá, Guatemala.

Marjanovic, M., Kovacevic, M., Bajat, B., Mihalic, S., & Abolmasov, B. (2011). Landslide Assessment of the Starca Basin (Croatia) using machine learning algorithms. *ACTA GEOTECHNICA SLOVENICA*, *8*(2), 45-55.

Miner, A. S., Vamplew, P., Windle, D. J., Flentje, P., & Warner, P. (2010). A comparative study of various data mining techniques as applied to the modeling of landslide susceptibility on the Bellarine Peninsula, Victoria, Australia.

Mujalli, R. O., & De ONa, J. (2011). A method for simplifying the analysis of traffic accidents injury severity on two-lane highways using Bayesian networks.*Journal of Safety Research*, *42*(5), 317-326.

Nandi, A., & Shakoor, A. (2010). A GIS-based landslide susceptibility evaluation using bivariate and multivariate statistical analyses. *Engineering Geology*, *110*(1), 11-20.

Newhall, C. G. (1987a). Geology of the Lake Atitlan region, western Guatemala. *Journal of Volcanology and Geothermal Research*, *33* (1), 23-55.

Newhall, C. G., Paull, C. K., Bradbury, J. P., Higuera-Gundy, A., Poppe, L. J., Selp'C, S., Bonar Sharpless, N., & Ziagos, J. (1987b). Recent geologic history of Lake Atitlan, a caldera lake in western Guatemala. *Journal of volcanology and geothermal research*, *33* (1), 81-107.

Oommen, T., Baise, L. G., & Vogel, R. (2010). Validation and application of empirical liquefaction models. *Journal of geotechnical and geoenvironmental engineering*, *136*(12), 1618-1633.

Ozdemir, A. (2011). Landslide susceptibility mapping using Bayesian approach in the Sultan Mountains (Akşehir, Turkey). *Natural hazards*, *59*(3), 1573-1607.

Pathak, P. Topographic Wetness Index. (2010). Python script for ArcGIS. http://arcscripts.esri.com/details.asp?dbid=16750. 10 October 2012.

Pourghasemi, H. R., Pradhan, B., Gokceoglu, C., Mohammadi, M., & Moradi, H. R. (2012). Application of weights-of-evidence and certainty factor models and their comparison in landslide susceptibility mapping at Haraz watershed, Iran.*Arabian Journal of Geosciences*, 1-15.

Pradhan, B., Oh, H.J. & Buchroithner, M. (2010): Weights of- evidence model applied to landslide susceptibility mapping in a tropical hilly area, Geomatics, Natural Hazards and Risk, 1:3, 199-223

Pradhan, B., & Youssef, A. M. (2010). Manifestation of remote sensing data and GIS on landslide hazard analysis using spatial-based statistical models. *Arabian Journal of Geosciences*, *3*(3), 319-326.

R Core Team. (2013). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. http://www.R-project.org

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.C., and Müller, M. (2012). pROC: display and analyze ROC curves. R Package.

Saha, A. K., Gupta, R. P., Sarkar, I., Arora, M. K., & Csaplovics, E. (2005). An approach for GIS-based statistical landslide susceptibility zonation—with a case study in the Himalayas. *Landslides*, *2*(1), 61-69.

Saito, H., Nakayama, D., & Matsuyama, H. (2009). Comparison of landslide susceptibility based on a decision-tree model and actual landslide occurrence: the Akaishi Mountains, Japan. *Geomorphology*, *109*(3), 108-121.

Segeplan (Secretary of Planning and Programming of the Presidency). (2006). Orhto-photos and Land use. http://www.segeplan.gob.gt/2.0/index.php?option=com_wrapper&view=wrapper&Itemid=260.

Stevens, D.L., and A.R. Olsen. 2004. "Spatially balanced sampling of natural resources."*Journal of the American Statistical Association* 99 (465): 262–278.

Silverman, J. (2011). *"The need obliged us": Culture as capacity during Hurricane Stan emergency response. A case study from Tectitán, Huehuetenango, Guatemala.* MS thesis, Michigan Technological University). Retrieved http://www.geo.mtu.edu/rs4hazards/Project%20resources/theses/silverman_thesis.pdf

Song, Y., Gong, J., Gao, S., Wang, D., Cui, T., Li, Y., & Wei, B. (2012). Susceptibility assessment of earthquake-induced landslides using Bayesian network: A case study in Beichuan, China. *Computers & Geosciences*, *42*, 189-199.

Süzen, M. L., & Doyuran, V. (2004). A comparison of the GIS based landslide susceptibility assessment methods: multivariate versus bivariate. *Environmental Geology*, *45*(5), 665-679.

Theobald, D.M., D.L. Stevens, Jr., D. White, N.S. Urquhart, A.R. Olsen, and J.B. Norman. 2007. "Using GIS to Generate Spatially Balanced Random Survey Designs for Natural Resource Applications."*Environmental Management* 40: 134–146.

Tien Bui, D., Pradhan, B., Lofman, O., & Revhaug, I. (2012). Landslide susceptibility assessment in Vietnam using support vector machines, decision tree, and Naive Bayes Models. *Mathematical Problems in Engineering*, *2012*.

Van Westen, C. J., Rengers, N., Terlien, M. T. J., & Soeters, R. (1997). Prediction of the occurrence of slope instability phenomenal through GIS-based hazard zonation. *Geologische Rundschau*, *86*(2), 404-414.

Van Westen, C. J. (2004). Geo-information tools for landslide risk assessment: an overview of recent developments. In *Landslides: Evaluation and Stabilization—Glissement de Terrain: Evaluation et Stabilisation: Proceedings of the 9th International Symposium on Landslides. Rio de Janeiro, Brazil* (pp. 39-56).

Van Westen, C. J., Castellanos, E., & Kuriakose, S. L. (2008). Spatial data for landslide susceptibility, hazard, and vulnerability assessment: an overview.*Engineering geology*, *102*(3), 112-131.

Witten, I. H., & Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques, Second Edition*. Morgan Kaufmann.

Wood, E. F., Sivapalan, M., & Beven, K. (1990). Similarity and scale in catchment storm response. *Reviews of Geophysics*, *28*(1), 1-18.

Yesilnacar E., Topal T. (2005); Landslide susceptibility mapping: A comparison of logistic regression and neural networks methods in a medium scale study, Hendek region (Turkey). *Engineering Geology,* 79 (3-4) , pp. 251-266.

Appendix 1: Watershed Geology and Geomorphology Map

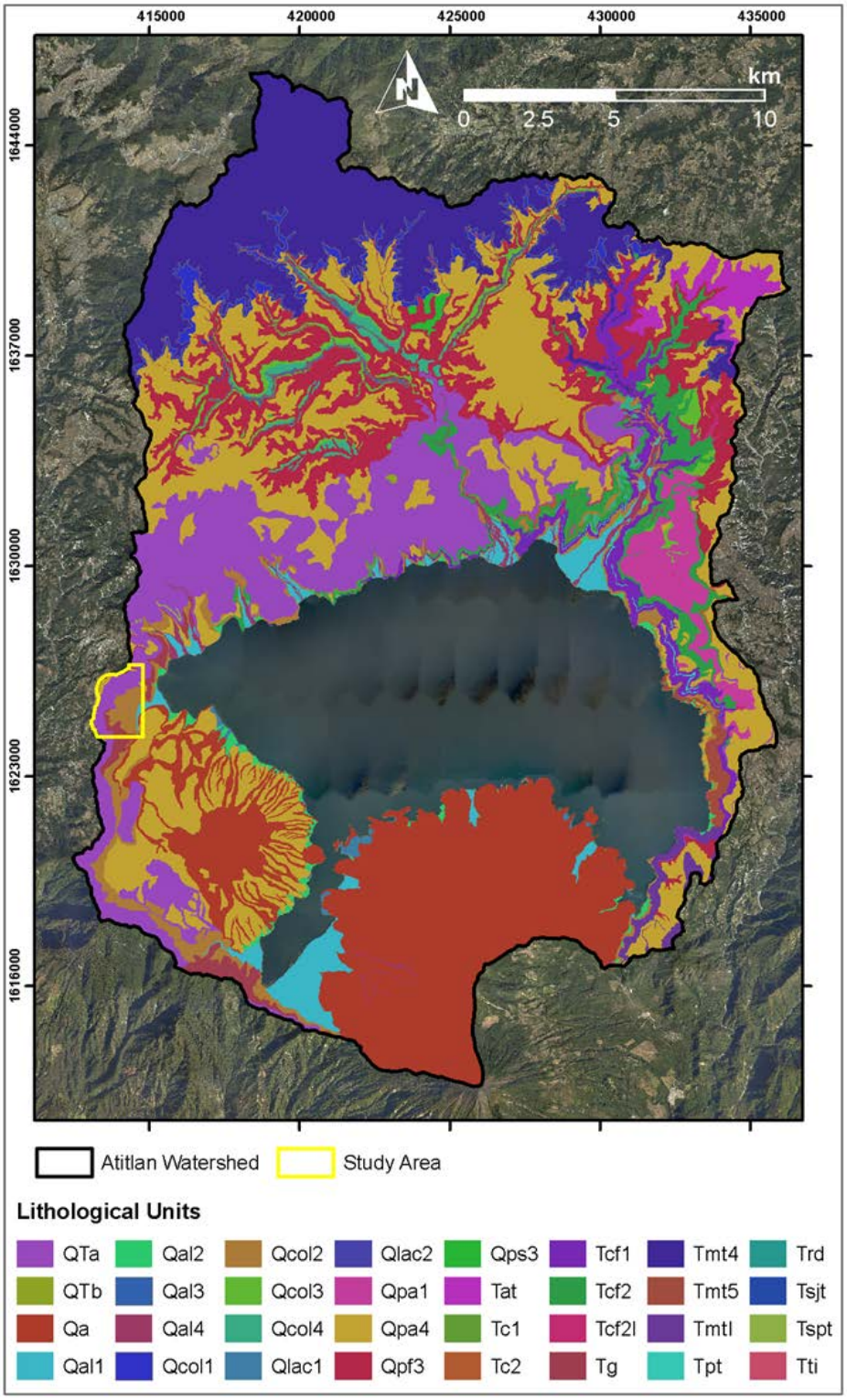


Figure 0.1: The Lake Atitlán geological map.

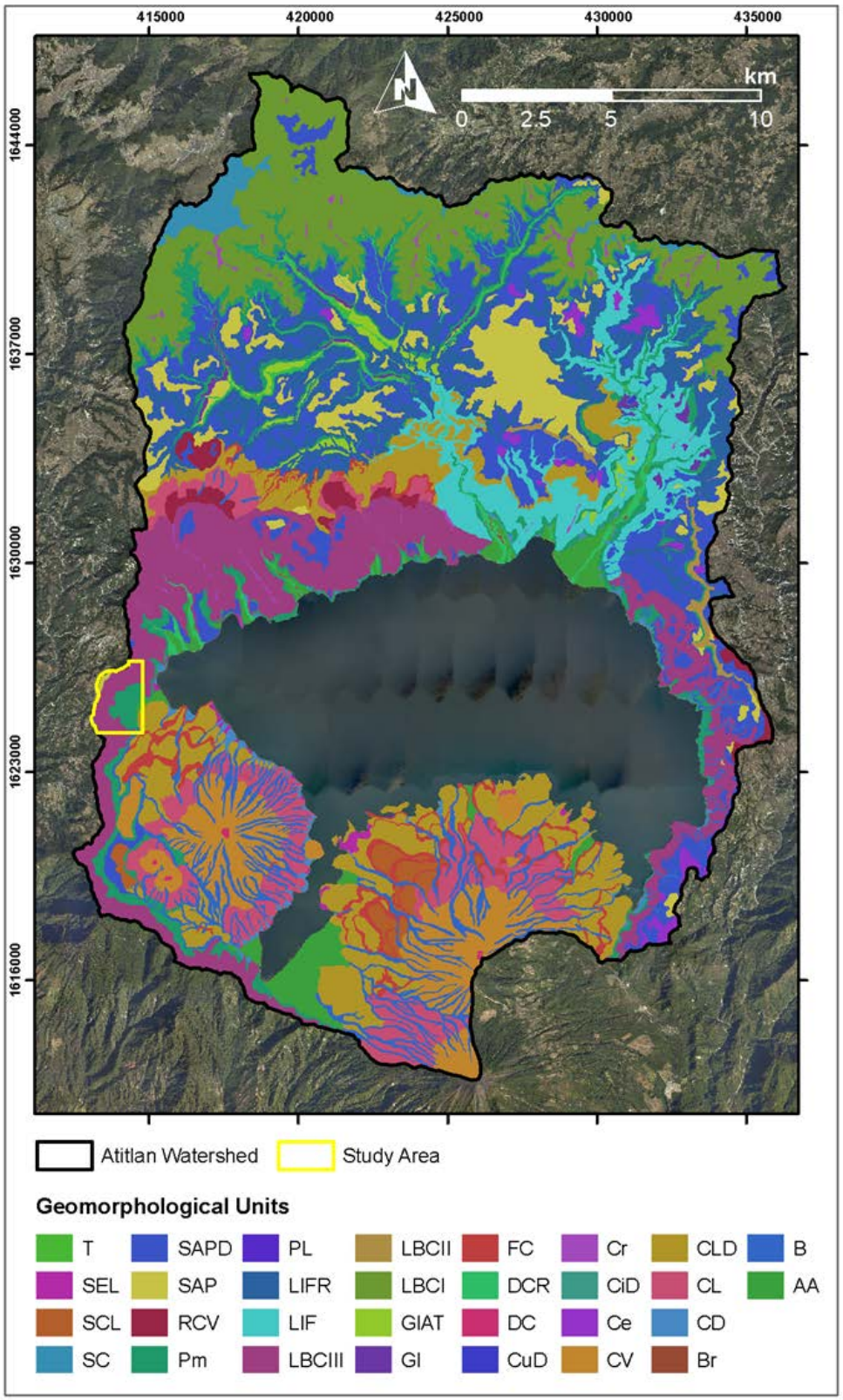


Figure 0.2: The Lake Atitlán geomorphology map.

Appendix 2: Attribute Data for each Point in Data Sets A and B

Appendix 2 is an electronic appendix and can be found in the attached excel file entitled *AttrData_NLS-LS_A-B.xlsx*. The excel workbook has two sheets, one called Data Set A and another called Data Set B.

ID	Dist. to Fault (m)	Geology (unit #)	Geo-Morph. (unit #)	Slope (degree)	Prof. Curv.	NLS (N) or LS (Y)	Bay-6 Prediction of NLS (N) or LS (Y)	Bay-6 Probability of NLS or LS (percent)
1	400	3	7	6.574	-0.015	N	Y	0.666
2	50	3	7	9.649	0.115	N	Y	0.926
3	50	2	8	11.502	-0.125	N	N	0
4	200	3	7	7.125	0.090	N	Y	0.926
5	300	5	3	28.303	-1.345	N	N	0
6	100	6	8	30.340	0.798	N	Y	0.926
7	400	6	8	43.237	-0.062	N	N	0
8	200	2	8	29.571	0.279	N	N	0
9	400	6	8	30.525	-2.912	N	N	0
10	200	7	5	4.076	0.063	N	N	0
11	50	3	8	10.135	0.091	N	N	0
12	200	6	8	26.683	0.196	N	N	0
13	200	6	8	46.095	1.073	N	N	0.371
14	100	6	8	22.927	-1.718	N	N	0.371
15	400	3	7	9.649	0.731	N	N	0
16	200	3	7	4.975	0.287	N	N	0
17	50	6	8	37.368	0.122	N	N	0.371
18	400	6	8	40.730	-0.123	N	Y	0.666
19	400	1	4	0.506	0.000	N	N	0
20	300	3	7	13.807	0.010	N	N	0.001
21	300	6	8	40.392	0.250	N	N	0.001
22	100	6	8	44.312	0.237	N	N	0.371
23	500	4	8	39.996	-0.160	N	N	0
24	500	1	4	2.725	0.000	N	N	0.075
25	50	3	7	7.877	0.259	N	N	0
26	100	6	8	39.761	0.173	N	N	0.371
27	50	6	8	31.509	0.059	N	N	0
28	200	6	8	29.327	0.444	N	N	0.371
29	400	5	2	1.432	0.250	N	N	0.371
30	300	3	7	10.422	0.268	N	N	0
31	50	5	5	4.319	-0.287	N	N	0

ID	Dist. to Fault (m)	Geology (unit #)	Geo-Morph. (unit #)	Slope (degree)	Prof. Curv.	NLS (N) or LS (Y)	Bay-6 Prediction of NLS (N) or LS (Y)	Bay-6 Probability of NLS or LS (percent)
32	100	3	7	12.614	0.140	N	N	0
33	200	6	8	45.890	0.100	N	N	0.001
34	100	3	7	11.617	-0.471	N	N	0.009
35	200	6	8	38.511	0.556	N	N	0.008
36	100	6	8	45.742	-0.807	N	N	0
37	50	6	8	30.679	0.051	N	N	0.371
38	300	2	8	31.106	-0.231	N	N	0.371
39	500	5	2	5.711	-0.500	N	N	0
40	300	3	7	8.775	0.038	N	N	0.371
41	200	2	8	27.142	0.260	N	Y	0.512
42	200	3	7	9.257	0.000	N	N	0.371
43	200	3	7	19.454	0.168	N	Y	0.666
44	300	5	2	11.906	-1.412	N	Y	0.926
45	50	3	7	13.642	0.048	N	N	0.371
46	100	6	8	40.895	-1.099	N	N	0.001
47	200	6	8	45.789	1.345	N	N	0
48	300	6	8	47.816	0.050	N	Y	0.666
49	200	1	7	3.949	0.250	N	N	0
50	100	6	8	33.305	0.525	N	N	0.371
51	200	6	8	48.844	-0.544	N	N	0.371
52	100	6	8	31.182	-0.423	N	N	0.371
53	400	3	7	3.949	0.250	N	N	0.371
54	200	3	7	7.213	-0.172	N	N	0
55	200	6	8	36.843	0.283	N	N	0
56	600	6	8	39.052	-0.371	N	Y	0.926
57	500	5	2	9.257	0.050	N	Y	0.666
58	600	3	7	11.701	0.050	N	N	0.371
59	50	3	7	13.986	0.050	N	N	0.371
60	300	3	7	25.230	0.838	N	N	0.371
61	400	5	2	2.725	0.000	N	Y	0.787
62	300	6	8	33.606	-0.182	N	N	0.237
63	400	6	8	41.486	-1.455	N	Y	0.666
64	100	6	8	45.532	0.180	N	N	0.371
65	300	3	7	6.096	0.044	N	Y	0.926
66	200	3	7	10.802	0.115	N	N	0.371
67	500	6	8	45.023	-0.012	N	Y	0.666

ID	Dist. to Fault (m)	Geology (unit #)	Geo-Morph. (unit #)	Slope (degree)	Prof. Curv.	NLS (N) or LS (Y)	Bay-6 Prediction of NLS (N) or LS (Y)	Bay-6 Probability of NLS or LS (percent)
68	100	2	8	25.211	1.333	N	Y	0.787
69	300	6	8	48.549	-0.936	N	Y	0.926
70	500	3	7	7.552	-0.207	N	Y	0.926
71	50	3	7	7.618	0.030	N	Y	0.926
72	200	6	8	40.377	0.199	N	Y	0.926
73	300	3	7	8.775	0.038	N	Y	0.666
74	50	3	7	29.651	0.562	N	Y	0.666
75	50	6	8	37.326	0.067	N	N	0.371
76	300	2	8	26.497	0.476	N	Y	0.666
77	400	5	2	7.213	-0.050	N	N	0
78	300	3	7	6.320	0.015	N	N	0
79	100	3	7	9.738	0.233	N	N	0
80	50	3	7	16.754	0.519	N	N	0
81	200	3	7	25.438	-0.263	N	N	0
82	200	5	2	16.808	0.050	N	N	0.434
83	300	3	7	9.649	0.115	N	N	0
84	100	6	8	28.060	0.264	N	N	0.434
85	200	6	8	38.252	0.000	N	N	0
86	100	6	8	32.537	0.096	N	N	0.27
87	200	2	8	25.758	0.218	N	Y	0.711
88	300	6	8	54.950	0.294	N	N	0.434
89	400	3	7	9.636	-0.233	N	Y	0.711
90	100	7	1	24.082	1.520	N	N	0
91	200	2	8	21.568	-0.167	N	N	0
92	200	6	8	35.283	-0.262	N	N	0
93	400	6	8	36.055	0.523	N	N	0.27
94	200	5	3	21.257	0.207	N	N	0
95	300	3	7	13.342	0.175	N	Y	0.711
96	100	3	7	14.303	-0.067	N	N	0.434
97	200	6	8	43.357	-0.104	N	N	0
98	500	6	8	43.457	0.472	N	N	0
99	500	5	2	4.289	0.250	N	N	0.27
100	400	3	7	2.530	0.500	N	N	0.001
101	200	2	8	33.901	-0.043	N	N	0
102	50	3	7	16.754	0.329	N	N	0
103	200	6	8	39.177	-0.719	N	N	0.434

ID	Dist. to Fault (m)	Geology (unit #)	Geo-Morph. (unit #)	Slope (degree)	Prof. Curv.	NLS (N) or LS (Y)	Bay-6 Prediction of NLS (N) or LS (Y)	Bay-6 Probability of NLS or LS (percent)
104	200	5	2	6.239	-0.231	N	N	0
105	100	6	8	39.490	0.761	N	N	0.434
106	500	6	8	33.352	-0.726	N	N	0.434
107	400	3	7	8.775	0.305	N	N	0.434
108	50	3	7	16.788	-0.826	N	N	0.009
109	100	2	8	23.834	0.540	N	N	0.434
110	100	6	8	54.415	0.650	N	N	0.27
111	400	6	8	48.680	0.297	N	Y	0.711
112	300	3	7	6.419	0.000	N	N	0.434
113	50	7	1	6.075	-0.044	N	N	0.434
114	300	3	7	6.574	0.015	N	N	0
115	100	3	7	18.602	0.353	N	N	0
116	400	6	8	39.603	-0.238	N	N	0.434
117	200	3	7	4.548	0.063	N	N	0
118	200	6	8	26.626	-0.921	N	N	0.434
119	100	6	8	34.453	0.363	N	N	0
120	400	2	8	28.972	0.490	N	N	0
121	300	6	8	44.224	0.172	N	Y	0.935
122	300	5	2	11.234	-0.899	N	Y	0.711
123	300	3	7	7.684	0.274	N	N	0.001
124	100	3	7	7.072	0.150	N	N	0.434
125	100	2	8	18.165	0.122	N	N	0.434
126	50	3	7	10.074	0.152	N	Y	0.705
127	300	6	8	36.267	0.125	N	Y	0.542
128	200	3	7	12.499	-0.206	N	N	0.434
129	100	6	8	27.352	0.518	N	Y	0.711
130	50	3	7	14.036	0.440	N	N	0.434
131	300	3	7	5.885	-0.074	N	N	0.001
132	50	3	7	11.586	0.102	N	N	0.434
133	100	6	8	38.308	-0.221	N	Y	0.817
134	100	6	8	23.317	-0.456	N	N	0.434
135	400	6	8	41.762	0.034	N	N	0.434
136	100	6	8	41.304	0.350	N	Y	0.711
137	500	3	7	5.644	-0.115	N	Y	0.874
138	50	3	7	20.244	0.250	N	Y	0.711
139	200	6	8	44.988	-0.813	N	Y	0.935

ID	Dist. to Fault (m)	Geology (unit #)	Geo-Morph. (unit #)	Slope (degree)	Prof. Curv.	NLS (N) or LS (Y)	Bay-6 Prediction of NLS (N) or LS (Y)	Bay-6 Probability of NLS or LS (percent)
140	200	6	8	44.660	-0.175	N	Y	0.817
141	500	6	8	48.374	0.750	N	Y	0.981
142	300	2	8	38.650	0.480	N	N	0.434
143	400	3	7	9.926	-0.250	N	Y	0.935
144	50	3	7	13.360	0.000	N	Y	0.935
145	200	5	8	24.628	-0.100	N	N	0.434
146	300	6	8	47.056	-0.372	N	Y	0.898
147	50	6	8	42.475	-1.516	N	Y	0.935
148	500	3	7	6.158	-0.044	N	Y	0.711
149	100	3	7	8.902	-0.150	N	Y	0.711
150	200	6	8	43.616	0.036	N	Y	0.542
151	300	2	8	38.252	0.293	N	N	0.434
152	400	5	2	6.055	-0.125	N	N	0
153	400	3	7	7.668	-0.216	N	N	0.003
154	50	3	7	10.037	0.438	N	N	0
155	200	3	7	11.834	0.000	N	N	0
156	50	3	7	13.387	0.255	N	Y	0.678
157	300	6	8	37.851	-0.300	N	N	0.256
158	200	3	7	7.552	-0.063	N	Y	0.544
159	50	3	8	21.191	1.421	N	N	0
160	200	2	8	28.514	-0.224	N	N	0
161	400	3	7	6.555	0.050	N	Y	0.81
162	50	3	7	7.125	0.090	N	N	0.378
163	50	3	7	12.067	0.191	N	N	0
164	50	3	7	10.219	0.038	N	Y	0.81
165	300	3	7	9.054	-0.595	N	N	0
166	200	6	8	30.027	0.161	N	Y	0.937
167	200	6	8	41.911	-0.010	N	N	0.378
168	200	6	8	39.510	0.118	N	N	0
169	400	6	8	35.340	0.551	N	N	0.378
170	300	2	8	19.926	0.388	N	N	0.378
171	300	6	8	42.404	-0.764	N	N	0.378
172	300	1	4	2.581	-0.050	N	N	0.118
173	50	2	8	17.708	0.366	N	N	0
174	300	6	8	32.482	0.750	N	N	0
175	100	6	8	34.355	0.543	N	Y	0.937

ID	Dist. to Fault (m)	Geology (unit #)	Geo-Morph. (unit #)	Slope (degree)	Prof. Curv.	NLS (N) or LS (Y)	Bay-6 Prediction of NLS (N) or LS (Y)	Bay-6 Probability of NLS or LS (percent)
176	200	6	8	42.303	0.279	N	N	0.003
177	300	3	7	6.574	-0.015	N	N	0.378
178	200	3	7	8.888	0.000	N	Y	0.708
179	100	6	8	32.049	1.139	N	Y	0.678
180	400	6	8	47.384	0.558	N	Y	0.678
181	500	1	4	1.601	0.250	N	Y	0.678
182	200	3	7	9.649	0.075	N	N	0
183	100	6	8	45.533	-0.250	N	N	0
184	500	5	3	29.787	-0.085	N	N	0.378
185	200	6	8	48.545	-0.395	N	N	0
186	400	3	7	2.530	0.000	N	N	0
187	50	3	7	13.431	0.498	N	N	0
188	50	6	8	42.423	-0.101	N	N	0
189	100	6	8	49.710	0.950	N	Y	0.937
190	200	5	5	3.200	0.050	N	N	0.378
191	400	5	2	8.775	-0.526	N	N	0.118
192	400	3	7	7.000	0.226	N	N	0.378
193	50	1	4	2.530	0.188	N	N	0.227
194	100	3	7	10.327	0.058	N	N	0
195	50	3	7	17.269	-0.245	N	N	0.256
196	400	6	2	9.040	-0.625	N	Y	0.544
197	300	6	8	46.095	0.330	N	N	0
198	200	3	7	8.603	0.091	N	N	0.378
199	200	6	8	31.549	-0.622	N	N	0
200	200	3	7	13.520	0.300	N	N	0
201	200	3	7	11.502	0.471	N	N	0.378
202	200	2	8	29.139	0.282	N	Y	0.937
203	50	3	7	14.580	0.240	N	N	0
204	200	3	7	22.337	0.039	N	N	0.378
205	300	5	2	4.045	-0.250	N	N	0.256
206	500	6	8	25.576	0.116	N	N	0.378
207	100	3	7	11.417	0.469	N	N	0.003
208	200	6	8	32.152	-0.587	N	Y	0.678
209	100	6	8	43.314	-0.021	N	N	0.378
210	200	6	8	34.854	0.436	N	Y	0.678
211	200	2	8	26.522	0.169	N	Y	0.937

ID	Dist. to Fault (m)	Geology (unit #)	Geo-Morph. (unit #)	Slope (degree)	Prof. Curv.	NLS (N) or LS (Y)	Bay-6 Prediction of NLS (N) or LS (Y)	Bay-6 Probability of NLS or LS (percent)
212	300	6	8	48.374	0.261	N	Y	0.81
213	200	1	4	4.520	-0.250	N	N	0.003
214	100	2	8	17.580	-0.244	N	Y	0.937
215	100	6	8	39.569	-0.674	N	Y	0.678
216	200	6	8	46.448	-0.129	N	N	0.378
217	50	6	8	31.311	0.108	N	Y	0.678
218	400	3	7	4.319	-0.287	N	N	0.378
219	300	3	7	6.555	-0.050	N	Y	0.678
220	100	6	8	34.994	-0.499	N	Y	0.937
221	600	6	8	40.157	-0.295	N	Y	0.937
222	500	5	2	11.997	-0.100	N	N	0
223	400	3	7	3.239	0.000	N	N	0.378
224	200	2	8	30.502	0.180	N	Y	0.678
225	100	3	7	13.986	0.869	N	N	0.256
226	300	6	8	33.094	-0.493	N	Y	0.937
227	300	5	2	2.581	-0.050	N	N	0.353
228	200	6	8	38.483	0.241	N	N	0
229	100	6	8	51.933	0.235	N	N	0.006
230	300	1	4	2.862	0.000	N	N	0
231	100	3	7	15.270	-0.089	N	N	0.353
232	50	2	8	25.164	0.691	N	N	0
233	200	6	8	39.094	0.324	N	N	0
234	500	6	8	41.395	-0.233	N	Y	0.92
235	50	2	8	19.920	0.442	N	Y	0.766
236	500	3	7	8.647	-0.493	N	N	0
237	100	2	8	14.036	0.240	N	N	0
238	100	3	7	12.077	-0.250	N	N	0
239	200	3	7	15.444	0.189	N	Y	0.766
240	400	6	8	43.967	-0.215	N	N	0
241	200	6	8	38.021	1.561	N	N	0
242	300	3	7	8.545	-0.038	N	Y	0.658
243	50	3	7	24.413	0.021	N	N	0
244	200	6	8	45.811	-0.063	N	N	0.001
245	400	2	8	33.967	0.264	N	N	0
246	500	5	2	8.127	-0.259	N	N	0.001
247	200	3	7	10.147	-0.250	N	N	0.353

ID	Dist. to Fault (m)	Geology (unit #)	Geo-Morph. (unit #)	Slope (degree)	Prof. Curv.	NLS (N) or LS (Y)	Bay-6 Prediction of NLS (N) or LS (Y)	Bay-6 Probability of NLS or LS (percent)
248	50	3	7	11.406	0.088	N	Y	0.658
249	100	3	7	7.956	-0.034	N	N	0
250	100	3	7	13.832	0.626	N	N	0.012
251	200	6	8	24.628	-0.500	N	N	0.07
252	500	4	8	37.039	-0.642	N	N	0
253	200	3	7	9.376	0.100	N	N	0.353
254	100	6	8	28.232	0.061	N	Y	0.766
255	200	6	8	31.686	-0.103	N	N	0
256	400	5	2	5.272	-0.813	N	N	0
257	300	3	7	8.545	-0.038	N	N	0
258	200	3	7	11.533	-0.188	N	N	0.238
259	200	6	8	40.133	-0.279	N	N	0
260	100	6	8	48.975	0.192	N	N	0.353
261	300	6	8	43.081	0.528	N	N	0
262	100	6	8	40.248	-0.088	N	Y	0.766
263	300	3	7	5.272	0.115	N	N	0
264	100	3	7	12.756	0.250	N	N	0.353
265	500	6	8	35.009	-1.690	N	Y	0.766
266	500	5	2	3.682	0.000	N	Y	0.658
267	100	3	7	12.344	-0.250	N	N	0
268	200	6	8	40.799	0.528	N	N	0
269	200	5	2	6.379	0.000	N	N	0
270	200	6	8	35.936	-0.818	N	N	0.353
271	400	6	8	42.805	0.246	N	N	0.353
272	50	6	8	47.271	0.265	N	N	0.353
273	400	3	7	4.319	0.287	N	N	0
274	50	3	7	9.649	0.000	N	N	0
275	100	6	8	48.119	-0.120	N	N	0
276	100	5	8	27.971	-0.278	N	N	0.353
277	200	6	8	43.358	-0.106	N	N	0
278	400	6	8	46.239	0.509	N	N	0.238
279	200	2	8	29.099	-0.288	N	N	0
280	300	5	2	10.982	-0.890	N	N	0
281	400	3	7	7.160	0.250	N	N	0
282	50	3	7	17.765	1.067	N	N	0
283	300	3	7	7.552	-0.226	N	Y	0.92

ID	Dist. to Fault (m)	Geology (unit #)	Geo-Morph. (unit #)	Slope (degree)	Prof. Curv.	NLS (N) or LS (Y)	Bay-6 Prediction of NLS (N) or LS (Y)	Bay-6 Probability of NLS or LS (percent)
284	500	6	8	41.406	-0.659	N	N	0.353
285	300	6	8	37.574	0.502	N	Y	0.766
286	200	3	7	4.319	-0.213	N	Y	0.92
287	100	3	7	12.017	-0.750	N	N	0.353
288	50	6	8	29.544	0.367	N	N	0.353
289	400	2	8	25.007	0.615	N	Y	0.766
290	300	6	8	41.756	0.467	N	N	0.353
291	300	6	2	26.247	-1.123	N	Y	0.92
292	200	3	7	11.278	-0.242	N	Y	0.658
293	200	3	7	8.603	-0.409	N	N	0.353
294	100	2	8	22.905	0.239	N	Y	0.658
295	50	3	7	12.499	-0.206	N	N	0.196
296	200	3	7	13.977	-0.180	N	Y	0.658
297	100	3	7	24.652	0.654	N	Y	0.766
298	200	6	8	34.526	-0.257	N	Y	0.92
299	50	3	7	9.520	0.250	N	N	0.353
300	200	6	8	37.039	0.400	N	Y	0.92
301	300	3	7	7.684	-0.274	N	N	0.353
302	200	6	8	37.828	0.291	N	Y	0.793
303	100	6	8	27.636	0.475	N	N	0
304	300	6	8	44.965	0.000	N	Y	0.667
305	500	3	7	10.025	-0.215	N	N	0.389
306	100	3	7	10.327	0.012	N	N	0
307	200	6	8	32.256	-0.347	N	N	0
308	500	6	8	45.020	0.006	N	N	0.236
309	600	5	2	4.520	-0.325	N	N	0.389
310	400	3	7	2.530	0.500	N	N	0.001
311	200	6	8	26.060	0.701	N	Y	0.667
312	200	6	8	46.430	0.001	N	N	0
313	50	6	8	44.546	0.294	N	N	0
314	500	3	7	5.461	0.015	N	N	0
315	100	3	7	14.104	-0.002	N	N	0.389
316	100	6	8	33.946	-0.017	N	N	0
317	50	6	8	39.782	-0.052	N	N	0
318	200	2	8	31.335	0.254	N	N	0
319	500	3	7	7.684	-0.226	N	N	0

ID	Dist. to Fault (m)	Geology (unit #)	Geo-Morph. (unit #)	Slope (degree)	Prof. Curv.	NLS (N) or LS (Y)	Bay-6 Prediction of NLS (N) or LS (Y)	Bay-6 Probability of NLS or LS (percent)
320	200	3	7	7.552	0.063	N	N	0.001
321	200	3	7	9.163	0.213	N	N	0
322	50	3	7	13.564	0.213	N	N	0.001
323	300	6	8	40.901	-0.830	N	Y	0.667
324	300	6	8	40.506	-0.833	N	N	0.001
325	200	6	8	31.010	-1.688	N	N	0
326	200	3	7	17.057	0.193	N	N	0.389
327	50	3	7	24.672	0.350	N	N	0
328	100	6	8	41.510	0.422	N	Y	0.923
329	200	3	7	14.112	0.180	N	N	0.001
330	500	5	2	6.726	-0.210	N	N	0.389
331	300	3	7	7.877	0.259	N	N	0.389
332	50	3	7	16.374	0.211	N	N	0
333	50	2	8	12.137	0.132	N	N	0.236
334	100	3	7	9.988	0.000	N	N	0
335	500	4	8	38.619	0.089	N	N	0
336	300	3	7	10.620	-0.398	N	Y	0.667
337	200	6	8	29.136	-0.105	N	N	0.389
338	300	6	8	47.860	-0.496	N	N	0.389
339	200	6	8	35.014	-0.219	N	N	0.389
340	300	2	8	32.118	1.209	N	N	0
341	200	7	5	3.200	0.050	N	Y	0.923
342	100	2	8	20.846	0.116	N	Y	0.793
343	50	3	8	11.015	0.459	N	Y	0.667
344	200	6	8	35.137	0.364	N	N	0.009
345	200	2	8	30.366	0.550	N	N	0.389
346	200	6	8	38.937	-0.327	N	Y	0.793
347	100	6	8	42.639	0.250	N	N	0.007
348	400	3	7	8.188	0.305	N	N	0
349	300	3	7	7.265	0.250	N	N	0
350	50	6	8	29.951	2.518	N	N	0.389
351	400	1	4	2.025	-0.250	N	Y	0.667
352	300	3	7	13.342	0.175	N	N	0.389
353	50	6	8	38.113	0.443	N	N	0
354	500	5	8	37.589	0.554	N	N	0
355	200	6	8	45.896	-0.177	N	N	0.389

ID	Dist. to Fault (m)	Geology (unit #)	Geo-Morph. (unit #)	Slope (degree)	Prof. Curv.	NLS (N) or LS (Y)	Bay-6 Prediction of NLS (N) or LS (Y)	Bay-6 Probability of NLS or LS (percent)
356	400	1	4	2.530	0.000	N	N	0.236
357	50	3	7	10.422	0.382	N	N	0
358	50	6	8	31.749	0.313	N	Y	0.667
359	400	6	8	34.078	-0.059	N	N	0.493
360	200	2	8	30.887	1.227	N	N	0.236
361	300	5	2	3.200	-0.350	N	Y	0.923
362	300	3	7	11.783	0.512	N	Y	0.667
363	50	5	5	3.949	-0.250	N	Y	0.923
364	100	3	7	9.889	-0.198	N	Y	0.667
365	50	3	7	16.317	0.311	N	Y	0.649
366	300	5	2	14.636	-0.438	N	Y	0.793
367	300	6	8	46.938	0.379	N	Y	0.667
368	200	3	7	7.072	0.150	N	Y	0.793
369	100	6	8	22.499	-1.098	N	Y	0.793
370	100	6	8	49.481	0.106	N	Y	0.667
371	100	6	8	35.804	0.097	N	Y	0.923
372	300	2	8	33.145	-0.443	N	Y	0.793
373	500	5	2	5.076	-0.063	N	Y	0.667
374	300	3	7	7.552	0.226	N	Y	0.667
375	50	3	7	12.604	0.100	N	Y	0.667
376	100	3	7	9.520	0.050	N	Y	0.793
377	200	3	7	19.544	-0.341	N	Y	0.696
378	300	6	8	42.862	-0.744	N	N	0.421
379	500	6	8	35.359	-0.461	N	N	0.001
380	100	3	7	8.589	-0.075	N	N	0
381	200	6	8	40.615	-0.453	N	N	0
382	200	6	8	42.713	0.426	N	N	0.421
383	300	6	8	34.559	0.269	N	N	0.421
384	100	5	5	5.576	-0.173	N	N	0.001
385	200	2	8	22.674	0.120	N	Y	0.822
386	200	6	8	34.513	-0.100	N	Y	0.936
387	400	5	2	1.432	-0.250	N	N	0.254
388	300	6	8	43.748	0.037	N	Y	0.696
389	500	3	7	5.553	0.188	N	N	0
390	200	3	7	6.320	-0.015	N	N	0.421
391	100	6	8	36.113	-0.500	N	N	0.421

ID	Dist. to Fault (m)	Geology (unit #)	Geo-Morph. (unit #)	Slope (degree)	Prof. Curv.	NLS (N) or LS (Y)	Bay-6 Prediction of NLS (N) or LS (Y)	Bay-6 Probability of NLS or LS (percent)
392	600	6	8	43.242	-0.667	N	N	0
393	500	5	2	7.569	-0.210	N	Y	0.696
394	500	3	7	2.950	0.000	N	N	0
395	200	2	8	28.489	0.038	N	N	0.001
396	200	3	7	9.585	0.250	N	Y	0.517
397	400	5	2	1.601	0.250	N	N	0.254
398	200	6	8	50.804	0.194	N	N	0.421
399	300	6	8	42.875	-1.254	N	N	0
400	100	6	8	42.667	0.655	N	N	0
401	400	1	4	2.530	0.000	N	Y	0.696
402	200	3	7	9.376	0.100	N	N	0.015
403	100	2	8	33.901	0.027	N	N	0
404	200	6	8	46.463	1.185	N	N	0
405	500	6	8	42.576	-0.250	N	N	0.421
406	100	2	8	19.572	-0.060	N	Y	0.696
407	200	6	8	45.277	0.293	N	N	0.254
408	500	3	7	5.711	0.000	N	N	0
409	50	3	7	20.979	-0.700	N	N	0
410	100	3	7	10.562	0.198	N	N	0.421
411	200	3	7	28.976	0.624	N	Y	0.822
412	300	6	8	47.444	0.138	N	N	0
413	200	6	8	45.472	-0.058	N	N	0
414	300	3	7	8.428	-0.195	N	N	0
415	50	3	7	28.332	0.404	N	N	0.421
416	50	6	8	42.785	0.007	N	N	0.254
417	300	2	8	38.303	-0.135	N	N	0.254
418	500	5	2	4.289	0.750	N	N	0.421
419	100	3	7	7.486	-0.172	N	Y	0.696
420	200	3	7	9.889	0.198	N	Y	0.696
421	50	3	7	12.315	0.618	N	N	0.001
422	100	3	7	12.604	0.000	N	N	0.009
423	200	5	2	17.510	-0.491	N	N	0.421
424	400	5	8	25.701	-0.560	N	Y	0.822
425	300	3	7	10.327	0.742	N	N	0.421
426	50	3	7	12.017	1.660	N	Y	0.696
427	100	6	8	33.188	-0.021	N	N	0.231

ID	Dist. to Fault (m)	Geology (unit #)	Geo-Morph. (unit #)	Slope (degree)	Prof. Curv.	NLS (N) or LS (Y)	Bay-6 Prediction of NLS (N) or LS (Y)	Bay-6 Probability of NLS or LS (percent)
428	100	2	8	23.936	0.288	N	N	0.254
429	200	2	8	26.392	0.121	N	N	0
430	300	3	7	4.520	0.250	N	N	0.001
431	200	3	8	14.410	0.240	N	Y	0.696
432	50	6	8	41.813	-0.154	N	Y	0.936
433	300	6	8	34.802	0.128	N	Y	0.696
434	400	3	7	9.216	0.463	N	N	0.421
435	50	3	7	18.997	0.291	N	Y	0.517
436	200	6	8	50.085	0.438	N	N	0.254
437	50	6	8	36.664	0.495	N	Y	0.936
438	500	6	8	40.852	0.257	N	Y	0.696
439	500	5	2	5.755	-0.100	N	Y	0.872
440	300	7	5	1.825	-0.250	N	Y	0.936
441	200	2	8	30.645	-0.166	N	Y	0.936
442	100	3	7	11.103	0.019	N	N	0.421
443	200	3	7	28.562	0.112	N	Y	0.696
444	200	5	2	6.158	0.074	N	Y	0.822
445	50	6	8	29.343	0.238	N	Y	0.936
446	400	6	8	40.627	0.098	N	Y	0.936
447	50	6	8	45.999	0.364	N	Y	0.936
448	400	3	7	6.574	-0.202	N	Y	0.517
449	100	6	8	44.364	-0.133	N	Y	0.696
450	300	6	8	48.714	0.131	N	N	0.254
451	200	2	8	19.942	-0.598	N	Y	0.936
452	400	3	7	7.435	-0.250	N	Y	0.501
453	50	5	5	7.265	0.346	N	N	0.25
454	200	3	7	7.000	-0.226	N	N	0
455	200	6	8	38.663	0.154	N	N	0
456	400	6	8	10.562	-0.897	N	N	0
457	200	3	7	5.644	0.050	N	N	0.393
458	100	6	8	31.643	0.390	N	N	0
459	50	6	8	35.266	1.305	N	N	0.393
460	400	6	8	37.833	0.777	N	Y	0.501
461	300	6	8	44.049	0.476	N	N	0
462	400	5	2	6.781	-0.250	N	N	0
463	200	3	7	8.428	-0.305	N	N	0.393

ID	Dist. to Fault (m)	Geology (unit #)	Geo-Morph. (unit #)	Slope (degree)	Prof. Curv.	NLS (N) or LS (Y)	Bay-6 Prediction of NLS (N) or LS (Y)	Bay-6 Probability of NLS or LS (percent)
464	100	3	7	11.048	-0.520	N	N	0
465	100	2	8	9.851	0.250	N	N	0.25
466	50	3	7	15.512	0.288	N	N	0.001
467	400	6	8	45.533	-1.000	N	N	0
468	100	3	7	12.137	0.368	N	N	0
469	100	3	7	9.468	-0.561	N	N	0.25
470	200	6	3	29.306	-1.966	N	N	0.393
471	200	6	8	33.221	0.555	N	N	0.393
472	200	6	8	37.770	-0.092	N	Y	0.926
473	50	3	7	18.780	0.193	N	N	0
474	300	6	8	54.461	1.009	N	N	0.001
475	300	3	7	6.055	0.000	N	Y	0.805
476	200	6	8	38.468	-0.063	N	N	0
477	50	6	8	33.276	0.523	N	N	0.25
478	400	6	8	42.744	-0.528	N	Y	0.805
479	100	6	8	43.166	0.451	N	N	0.25
480	500	3	7	6.379	0.000	N	Y	0.661
481	50	3	7	15.888	0.264	N	N	0
482	200	6	8	44.458	0.250	N	N	0
483	100	6	8	34.621	0.455	N	N	0
484	500	6	8	42.363	-0.183	N	N	0.393
485	200	2	8	22.747	0.622	N	N	0.393
486	300	3	7	1.601	0.250	N	N	0
487	200	3	7	12.604	0.100	N	N	0.001
488	100	5	8	30.525	-0.630	N	Y	0.661
489	200	5	2	8.545	0.112	N	N	0
490	200	6	8	43.690	-0.821	N	N	0
491	50	5	3	8.675	-0.055	N	N	0.08
492	500	3	7	2.950	-0.050	N	Y	0.661
493	50	3	7	16.310	0.215	N	N	0.393
494	50	6	8	34.121	0.548	N	N	0.393
495	50	6	8	51.058	1.063	N	Y	0.661
496	300	6	8	47.247	0.231	N	N	0.001
497	300	2	8	21.845	0.231	N	N	0.001
498	400	5	2	9.585	-0.640	N	N	0.393
499	500	3	7	6.574	-0.250	N	N	0.25

ID	Dist. to Fault (m)	Geology (unit #)	Geo-Morph. (unit #)	Slope (degree)	Prof. Curv.	NLS (N) or LS (Y)	Bay-6 Prediction of NLS (N) or LS (Y)	Bay-6 Probability of NLS or LS (percent)
500	50	3	7	12.470	0.144	N	N	0.393
501	200	3	7	8.902	-0.250	N	N	0.393
502	50	3	7	20.887	0.542	N	N	0.393
503	300	6	8	47.373	0.342	N	N	0.393
504	200	3	7	15.444	0.378	N	N	0
505	50	6	8	27.706	-1.250	N	N	0.25
506	200	2	8	19.336	1.428	N	N	0.393
507	400	3	7	6.855	-0.226	N	Y	0.805
508	50	3	7	12.765	0.510	N	Y	0.661
509	50	3	7	9.095	0.000	N	Y	0.926
510	600	6	8	40.730	0.227	N	N	0.393
511	300	3	7	13.387	-0.255	N	Y	0.501
512	100	6	8	33.111	-0.240	N	Y	0.926
513	300	5	3	33.755	-0.796	N	Y	0.926
514	100	3	7	22.041	0.000	N	Y	0.926
515	200	2	8	25.454	0.093	N	Y	0.661
516	300	6	8	45.831	0.265	N	N	0.393
517	300	7	5	1.132	0.250	N	Y	0.661
518	200	6	8	33.693	0.404	N	Y	0.661
519	50	6	8	29.306	0.898	N	Y	0.661
520	100	6	8	34.839	-0.004	N	Y	0.501
521	300	3	7	13.807	0.490	N	Y	0.926
522	100	3	7	13.387	0.154	N	Y	0.661
523	100	6	8	33.700	-0.393	N	N	0.25
524	500	1	4	1.132	-0.250	N	N	0.25
525	200	3	7	10.620	0.250	N	Y	0.661
526	300	6	8	32.545	0.849	N	Y	0.661
527	50	6	8	48.015	0.083	N	N	0
528	600	6	8	45.462	0.312	N	N	0.415
529	100	6	8	39.666	-1.804	N	N	0.415
530	400	1	4	1.432	0.250	N	N	0.415
531	100	3	7	15.354	0.660	N	N	0.415
532	50	6	8	40.272	-0.082	N	Y	0.808
533	100	6	8	45.650	0.896	N	N	0.009
534	500	6	8	50.607	-0.720	N	N	0
535	300	6	8	36.267	-0.504	N	Y	0.808

ID	Dist. to Fault (m)	Geology (unit #)	Geo-Morph. (unit #)	Slope (degree)	Prof. Curv.	NLS (N) or LS (Y)	Bay-6 Prediction of NLS (N) or LS (Y)	Bay-6 Probability of NLS or LS (percent)
536	400	5	2	3.393	-0.100	N	N	0.273
537	400	3	7	7.125	0.250	N	N	0.255
538	100	3	7	4.045	0.125	N	N	0.071
539	50	3	7	8.249	-0.014	N	N	0
540	100	3	7	11.533	0.164	N	N	0
541	300	6	8	11.103	-0.803	N	N	0
542	300	6	8	43.278	-0.461	N	N	0
543	200	3	7	17.529	0.150	N	Y	0.693
544	100	6	8	29.163	-0.855	N	N	0.415
545	200	5	3	30.021	-1.594	N	Y	0.522
546	200	3	7	11.331	-0.004	N	N	0
547	100	3	7	17.810	0.104	N	N	0
548	100	3	7	19.403	-0.100	N	Y	0.693
549	200	6	8	45.117	-0.033	N	Y	0.693
550	500	6	8	38.959	-0.099	N	N	0
551	200	3	7	10.147	-0.250	N	N	0
552	100	6	8	45.018	0.119	N	N	0
553	200	6	8	33.774	-0.014	N	N	0.01
554	200	2	8	25.750	0.051	N	N	0.001
555	200	6	8	41.042	0.714	N	N	0
556	200	7	5	5.576	0.250	N	N	0
557	50	2	8	10.422	0.198	N	Y	0.693
558	300	6	8	47.236	0.030	N	N	0
559	100	6	8	21.955	-0.538	N	N	0.415
560	400	3	7	3.682	0.000	N	N	0
561	500	4	8	33.411	-0.600	Y	N	0
562	400	6	8	33.444	-1.758	Y	N	0.415
563	400	6	8	33.444	-1.758	Y	N	0
564	100	6	8	50.655	0.063	Y	Y	0.693
565	100	6	8	49.833	-0.562	Y	Y	0.693
566	50	6	8	42.438	0.003	Y	N	0.415
567	500	6	8	44.100	-0.762	Y	Y	0.808
568	500	6	8	44.100	-0.762	Y	Y	0.808
569	200	6	8	46.561	-0.747	Y	N	0
570	300	2	8	33.525	-0.500	Y	N	0
571	300	2	8	30.045	-0.090	Y	N	0

ID	Dist. to Fault (m)	Geology (unit #)	Geo-Morph. (unit #)	Slope (degree)	Prof. Curv.	NLS (N) or LS (Y)	Bay-6 Prediction of NLS (N) or LS (Y)	Bay-6 Probability of NLS or LS (percent)
572	300	2	8	30.045	-0.090	Y	N	0
573	300	2	8	33.822	-0.233	Y	N	0
574	300	2	8	29.992	-1.180	Y	N	0
575	200	6	8	26.867	-1.716	Y	N	0
576	500	6	8	43.593	-0.535	Y	Y	0.693
577	500	6	8	43.593	-0.535	Y	N	0
578	500	4	8	26.465	-1.102	Y	N	0.415
579	500	4	8	33.411	-0.600	Y	N	0.415
580	500	4	8	37.964	-0.225	Y	Y	0.693
581	300	6	8	17.157	-0.941	Y	N	0.255
582	100	6	8	45.153	-0.011	Y	N	0
583	50	6	8	45.277	-0.564	Y	Y	0.93
584	400	6	8	50.225	-1.805	Y	N	0.196
585	400	6	8	50.225	-1.805	Y	Y	0.522
586	400	6	8	51.420	-2.028	Y	Y	0.93
587	400	6	8	51.420	-2.028	Y	Y	0.982
588	300	6	8	56.802	-0.154	Y	Y	0.693
589	300	6	8	56.802	-0.154	Y	N	0.415
590	300	6	8	56.802	-0.154	Y	Y	0.808
591	400	6	8	44.333	-1.834	Y	N	0.415
592	50	6	8	39.883	0.533	Y	Y	0.93
593	200	6	8	34.119	-0.203	Y	Y	0.693
594	400	6	8	20.624	-1.827	Y	N	0.196
595	400	6	8	20.624	-1.827	Y	Y	0.693
596	400	6	8	5.907	-0.350	Y	N	0.415
597	400	6	8	9.040	-0.625	Y	Y	0.808
598	300	6	2	8.675	-0.971	Y	N	0.415
599	300	6	8	26.853	-1.466	Y	Y	0.93
600	400	6	8	36.267	-1.667	Y	Y	0.808
601	400	6	8	36.267	-1.667	Y	Y	0.93
602	300	6	8	23.025	-1.603	Y	Y	0.679
603	300	6	8	33.104	-1.111	Y	Y	0.679
604	300	6	8	23.025	-1.603	Y	N	0.001
605	300	6	8	41.110	0.070	Y	N	0
606	300	6	8	41.110	0.070	Y	N	0.001
607	600	6	8	38.198	-0.198	Y	Y	0.541

ID	Dist. to Fault (m)	Geology (unit #)	Geo-Morph. (unit #)	Slope (degree)	Prof. Curv.	NLS (N) or LS (Y)	Bay-6 Prediction of NLS (N) or LS (Y)	Bay-6 Probability of NLS or LS (percent)
608	600	6	8	38.198	-0.198	Y	N	0.001
609	300	6	8	37.242	-1.568	Y	Y	0.541
610	300	6	8	37.242	-1.568	Y	N	0
611	500	4	8	34.133	-0.871	Y	N	0
612	500	4	8	37.921	0.048	Y	Y	0.679
613	500	4	8	36.217	-0.080	Y	N	0.401
614	400	6	8	41.990	0.432	Y	N	0.401
615	400	6	8	41.990	0.432	Y	Y	0.923
616	400	6	8	41.990	-0.499	Y	N	0
617	400	6	8	37.340	0.009	Y	N	0.401
618	500	6	8	46.410	0.489	Y	N	0
619	500	6	8	46.410	0.489	Y	N	0.401
620	500	6	8	45.895	0.483	Y	N	0
621	500	6	8	42.655	-0.334	Y	N	0.008
622	300	6	8	34.453	-1.254	Y	N	0
623	300	6	8	34.453	-1.254	Y	N	0.401
624	200	2	8	36.113	-0.622	Y	N	0
625	300	6	8	43.738	-0.491	Y	N	0
626	200	6	8	47.171	-0.101	Y	N	0.272
627	300	6	8	44.573	-0.765	Y	N	0
628	200	6	8	44.573	-0.765	Y	N	0
629	200	6	8	44.573	-0.765	Y	N	0.001
630	200	6	8	46.451	-0.316	Y	N	0
631	200	6	8	47.393	-0.125	Y	Y	0.679
632	200	6	8	46.927	0.316	Y	N	0
633	200	6	8	46.927	0.316	Y	N	0.401
634	200	6	8	32.026	-0.232	Y	N	0.401
635	200	6	8	36.839	-0.728	Y	N	0
636	200	6	8	32.026	-0.232	Y	N	0
637	300	6	8	35.675	-1.956	Y	Y	0.679
638	300	6	8	29.139	-1.140	Y	Y	0.679
639	300	2	8	27.410	-0.752	Y	Y	0.679
640	300	2	8	32.597	-1.054	Y	N	0
641	300	6	8	43.278	-0.461	Y	Y	0.679
642	300	6	8	43.278	-0.461	Y	N	0
643	200	6	8	35.691	-0.514	Y	Y	0.923

ID	Dist. to Fault (m)	Geology (unit #)	Geo-Morph. (unit #)	Slope (degree)	Prof. Curv.	NLS (N) or LS (Y)	Bay-6 Prediction of NLS (N) or LS (Y)	Bay-6 Probability of NLS or LS (percent)
644	50	6	8	31.098	0.007	Y	N	0
645	50	6	8	31.588	0.150	Y	N	0.01
646	100	6	8	34.684	0.298	Y	N	0
647	100	6	8	35.409	-0.243	Y	N	0.401
648	100	2	8	32.415	0.539	Y	N	0
649	200	6	8	44.425	0.561	Y	Y	0.679
650	200	6	8	44.425	0.561	Y	N	0
651	400	6	8	44.486	0.217	Y	N	0.001
652	400	6	8	45.882	0.259	Y	N	0.272
653	500	6	8	45.009	-0.500	Y	N	0
654	100	6	8	44.973	0.543	Y	N	0
655	300	2	8	29.808	0.290	Y	Y	0.541
656	300	2	8	31.593	0.122	Y	N	0.401
657	200	2	8	27.942	-0.188	Y	N	0
658	200	6	8	29.878	0.210	Y	Y	0.923
659	200	2	8	25.568	0.225	Y	Y	0.679
660	100	6	8	44.739	-0.722	Y	N	0.401
661	100	6	8	44.739	-0.722	Y	N	0.272
662	400	6	8	44.712	-0.195	Y	Y	0.923
663	400	6	8	44.712	-0.195	Y	N	0.001
664	400	6	8	44.712	-0.195	Y	N	0.272
665	400	5	8	17.197	-0.233	Y	Y	0.923
666	300	5	8	17.197	-0.233	Y	Y	0.679
667	300	6	8	21.376	-0.950	Y	Y	0.923
668	400	6	8	21.376	-0.950	Y	Y	0.923
669	400	6	8	21.376	-0.950	Y	Y	0.679
670	400	6	8	21.376	-0.950	Y	N	0.401
671	100	6	8	29.595	-0.918	Y	Y	0.923
672	100	6	8	29.595	-0.918	Y	Y	0.923
673	400	6	8	39.447	-0.731	Y	Y	0.923
674	400	6	8	39.447	-0.731	Y	Y	0.679
675	600	4	8	39.728	-0.119	Y	Y	0.679
676	400	6	8	41.824	-0.129	Y	Y	0.923
677	400	6	8	41.824	-0.129	Y	N	0.289
678	400	6	8	32.518	-0.250	Y	N	0
679	400	6	8	32.518	-0.250	Y	Y	0.699

ID	Dist. to Fault (m)	Geology (unit #)	Geo-Morph. (unit #)	Slope (degree)	Prof. Curv.	NLS (N) or LS (Y)	Bay-6 Prediction of NLS (N) or LS (Y)	Bay-6 Probability of NLS or LS (percent)
680	400	6	8	32.518	-0.250	Y	N	0
681	400	6	8	29.922	-0.250	Y	N	0.436
682	300	2	8	38.468	-0.464	Y	N	0
683	500	5	8	37.589	0.554	Y	N	0
684	500	4	8	44.360	-1.183	Y	N	0
685	50	6	8	46.334	0.088	Y	N	0
686	50	6	8	46.334	0.088	Y	N	0
687	50	6	8	42.438	0.003	Y	N	0.436
688	400	6	8	51.847	-1.713	Y	N	0.289
689	400	6	8	40.561	-2.489	Y	N	0
690	400	6	8	41.927	-2.269	Y	Y	0.816
691	400	6	8	54.152	-0.760	Y	N	0
692	400	6	8	43.576	-2.032	Y	Y	0.699
693	400	6	8	43.576	-2.032	Y	N	0.436
694	400	6	8	43.576	-2.032	Y	N	0
695	400	6	8	51.651	-0.975	Y	N	0.289
696	50	5	8	32.039	0.383	Y	N	0
697	400	6	8	45.160	0.648	Y	Y	0.816
698	400	6	8	45.160	0.648	Y	N	0
699	400	6	8	45.160	0.648	Y	N	0
700	400	6	8	50.048	0.188	Y	Y	0.549
701	200	6	8	30.295	-0.428	Y	N	0
702	300	6	8	35.359	-0.698	Y	Y	0.699
703	50	3	7	32.696	0.984	Y	Y	0.93
704	100	6	8	20.810	-0.007	Y	N	0
705	50	2	8	32.312	0.227	Y	Y	0.816
706	50	2	8	31.509	0.138	Y	N	0.013
707	400	6	8	45.109	-0.990	Y	N	0
708	400	6	8	45.109	-0.990	Y	N	0.436
709	200	6	8	15.391	-0.949	Y	N	0.289
710	300	6	2	14.636	-0.438	Y	N	0
711	300	6	8	14.636	-0.438	Y	N	0.436
712	300	6	8	35.485	0.900	Y	N	0
713	200	2	8	26.015	0.680	Y	Y	0.699
714	100	6	8	33.928	-0.597	Y	N	0
715	100	6	8	33.928	-0.597	Y	N	0

ID	Dist. to Fault (m)	Geology (unit #)	Geo-Morph. (unit #)	Slope (degree)	Prof. Curv.	NLS (N) or LS (Y)	Bay-6 Prediction of NLS (N) or LS (Y)	Bay-6 Probability of NLS or LS (percent)
716	300	6	8	45.222	-0.513	Y	N	0
717	300	6	8	45.222	-0.513	Y	N	0
718	500	6	8	35.099	-0.195	Y	N	0.436
719	500	6	8	34.526	-0.553	Y	N	0
720	400	5	8	19.301	-0.250	Y	N	0
721	300	6	8	45.999	-0.051	Y	N	0.001
722	300	6	8	45.077	0.236	Y	Y	0.699
723	400	6	8	37.574	0.277	Y	Y	0.699
724	600	6	8	35.468	0.458	Y	N	0
725	50	6	8	48.967	-0.159	Y	N	0.25
726	200	6	8	40.390	0.562	Y	N	0
727	200	6	8	38.883	0.132	Y	Y	0.816
728	200	6	8	42.109	-0.363	Y	N	0
729	400	6	8	45.000	0.000	Y	N	0.436
730	500	4	8	26.465	-1.102	Y	N	0
731	200	6	8	40.883	-0.698	Y	Y	0.93
732	300	6	8	46.028	1.025	Y	N	0.227
733	100	6	8	37.676	0.420	Y	Y	0.699
734	200	2	8	24.045	-0.264	Y	N	0.436
735	300	6	8	33.994	-0.021	Y	Y	0.93
736	500	6	8	45.365	0.250	Y	Y	0.98
737	300	6	8	29.787	-0.096	Y	Y	0.93
738	200	6	8	30.925	0.227	Y	N	0.1
739	400	6	8	40.378	-0.777	Y	Y	0.875
740	300	6	8	50.060	-0.632	Y	Y	0.895
741	600	6	8	40.175	-0.294	Y	N	0.072
742	400	6	8	24.833	-1.765	Y	Y	0.93
743	400	6	8	52.311	-0.076	Y	Y	0.93
744	400	6	8	45.250	-1.720	Y	Y	0.93
745	200	6	8	46.129	0.019	Y	N	0.436
746	200	2	8	26.392	0.121	Y	Y	0.93
747	200	2	8	26.392	0.121	Y	N	0.072
748	100	6	8	48.985	0.134	Y	Y	0.816
749	50	6	8	32.978	-2.599	Y	Y	0.699
750	100	6	8	29.034	-1.664	Y	Y	0.93
751	200	6	8	43.909	-0.152	Y	Y	0.93

Appendix 3: Permissions

SPRINGER LICENSE
TERMS AND CONDITIONS

Jul 18, 2013

This is a License Agreement between Patrice Cobin ("You") and Springer ("Springer") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Springer, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	3192080086411
License date	Jul 18, 2013
Licensed content publisher	Springer
Licensed content publication	Geologische Rundschau
Licensed content title	Prediction of the occurrence of slope instability phenomenal through GIS-based hazard zonation
Licensed content author	C. J. van Westen
Licensed content date	Jan 1, 1997
Volume number	86
Issue number	2
Type of Use	Thesis/Dissertation
Portion	Figures
Author of this Springer article	No
Order reference number	Fig. 2 and Fig. 4 in the paper.
Title of your thesis / dissertation	Probabilistic modeling of rainfall induced landslide hazard assessment in San Juan La Laguna, Sololá, Guatemala
Expected completion date	Aug 2013
Estimated size(pages)	91
Total	0.00 USD
Terms and Conditions	

Introduction
The publisher for this copyrighted material is Springer Science + Business Media. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions

apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

Limited License
With reference to your request to reprint in your thesis material on which Springer Science and Business Media control the copyright, permission is granted, free of charge, for the use indicated in your enquiry.

Licenses are for one-time use only with a maximum distribution equal to the number that you identified in the licensing process.

This License includes use in an electronic form, provided its password protected or on the university's intranet or repository, including UMI (according to the definition at the Sherpa website: http://www.sherpa.ac.uk/romeo/). For any other electronic use, please contact Springer at (permissions.dordrecht@springer.com or permissions.heidelberg@springer.com).

The material can only be used for the purpose of defending your thesis, and with a maximum of 100 extra copies in paper.

Although Springer holds copyright to the material and is entitled to negotiate on rights, this license is only valid, subject to a courtesy information to the author (address is given with the article/chapter) and provided it concerns original material which does not carry references to other sources (if material in question appears with credit to another source, authorization from that source is required as well).

Permission free of charge on this occasion does not prejudice any rights we might have to charge for reproduction of our copyrighted material in the future.

Altering/Modifying Material: Not Permitted
You may not alter or modify the material in any manner. Abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of the author(s) and/or Springer Science + Business Media. (Please contact Springer at (permissions.dordrecht@springer.com or permissions.heidelberg@springer.com)

Reservation of Rights
Springer Science + Business Media reserves all rights not specifically granted in the combination of (i) the license details provided by you and

accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

Copyright Notice:Disclaimer
You must include the following copyright and permission notice in connection with any reproduction of the licensed material: "Springer and the original publisher /journal title, volume, year of publication, page, chapter/article title, name(s) of author(s), figure number(s), original copyright notice) is given to the publication in which the material was originally published, by adding; with kind permission from Springer Science and Business Media"

Warranties: None

Example 1: Springer Science + Business Media makes no representations or warranties with respect to the licensed material.

Example 2: Springer Science + Business Media makes no representations or warranties with respect to the licensed material and adopts on its own behalf the limitations and disclaimers established by CCC on its behalf in its Billing and Payment terms and conditions for this licensing transaction.

Indemnity
You hereby indemnify and agree to hold harmless Springer Science + Business Media and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

No Transfer of License
This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without Springer Science + Business Media's written permission.

No Amendment Except in Writing
This license may not be amended except in a writing signed by both parties (or, in the case of Springer Science + Business Media, by CCC on Springer Science + Business Media's behalf).

Objection to Contrary Terms
Springer Science + Business Media hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and

conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and Springer Science + Business Media (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

Jurisdiction
All disputes that may arise in connection with this present License, or the breach thereof, shall be settled exclusively by arbitration, to be held in The Netherlands, in accordance with Dutch law, and to be conducted under the Rules of the 'Netherlands Arbitrage Instituut' (Netherlands Institute of Arbitration).***OR:***

All disputes that may arise in connection with this present License, or the breach thereof, shall be settled exclusively by arbitration, to be held in the Federal Republic of Germany, in accordance with German law.

Other terms and conditions:

v1.3

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number RLNK501069439.
Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.

Make Payment To:
Copyright Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006

For suggestions or comments regarding this order, contact RightsLink Customer Support:customercare@copyright.com or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.

Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.
